DOCUMENT RESUME

ED 082 838 PS 006 860

AUTHOR	Featherstone, Helen J.
TITLE	Cognitive Effects of Preschool Programs on Different Types of Children.
INSTITUTION	Huron Inst., Cambridge, Mass.
SPONS AGENCY	Office of Child Development (DHEW), Washington, D.C.
REPORT NO	OCD-H-1926
PUB DATE	Aug 73
NOTE	210p.
EDRS PRICE	MF-$0.65 HC-$9.87
DESCRIPTORS	*Achievement; Age Differences; *Educationally Disadvantaged; *Individual Differences; Literature Reviews; *Preschool Children; *Preschool Programs; Response Style (Tests); Sex Differences; Socioeconomic Status
IDENTIFIERS	*Planned Variations Study; Preschool Inventory; Stanford Binet

ABSTRACT

This study investigated whether various kinds of preschool programs have differential cognitive effects on different kinds of children. Relevant literature was reviewed and data, generated in the first 2 years of the Head Start Planned Variations Study (PVS), were analyzed. The eight preschool programs associated with the PVS were considered. Children's characteristics used for analysis were initial ability, previous school experience, sex, age, socioeconomic status, ethnicity, and style of response to testing. Cognitive outcomes were assessed with the Preschool Inventory and Stanford-Binet. Several tentative conclusions were drawn. First, some children's characteristics interact more powerfully than others with characteristics of preschool models. Response style and previous schooling seemed the most important variables. Second, where interactions are consistent, they follow suggestive patterns. Educationally disadvantaged children seem to achieve more in highly directive programs than in less directive ones. Third, the interactions may be significant factors in preschool success or failure. Many educational approaches may be needed, to be delivered in a flexible classroom designed to accomodate all kinds of children. (DP)

ED 082838

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

COGNITIVE EFFECTS OF PRESCHOOL PROGRAMS ON DIFFERENT TYPES OF CHILDREN

HELEN J. FEATHERSTONE

AUGUST, 1973

HURON INSTITUTE
CAMBRIDGE, MASSACHUSETTS

This document was prepared for Grant # H 1926 from the Office of Child Development, Department of Health, Education and Welfare, U. S. Government. The conclusions and recommendations in this report are those of the grantee and do not necessarily reflect the views of any federal agency.

The project director was Marshall S. Smith.

Project staff included:

Mary Jo Bane
Barbara Behrendt
Anthony Bryk
John Butler
Thomas Cerva
David Cohen
Jane David
Richard Elmore
Nathan Fox
David Gordon
Deborah Gordon
Sharon Hauck
Carol Lukas
Robert McMeekin
Anne Monaghan
David Napior
Ann Taylor
Deborah Walker
Herbert Weisberg
Jack Wiggins
Cicero Wilson
Cynthia Wohlleb
Joy Wolfe
Stanley Yutkins
Diane Zipperman

PS 006860

FILMED FROM BEST AVAILABLE COPY

TABLE OF CONTENTS

Introduction 1

Part I: Hypotheses
Literature Review and Report on the 1969-70 Planned Variation Data 15

Introduction to Part I 15
I. Initial Ability 19
Previous Research 22
69-70 Planned Variations Data 28
Hypotheses 35

II. Previous Preschool Experience 36
Previous Research 38
69-70 Planned Variations Data 45
Hypotheses 49

III. Sex 51
Previous Research 51
69-70 Planned Variations Data 53
Hypotheses 57

IV. Age 59
Previous Research 59
69-70 Planned Variations Data 61
Hypotheses 69

V. Socio-Economic Status 70
Previous Research 70
69-70 Planned Variations Data 71
Hypotheses 75

VI. Ethnicity 76
Previous Research 76
69-70 Planned Variations Data 77
Hypotheses 79

VII. Response Style 80
Previous Research 83
69-70 Planned Variations Data 85
Hypotheses 94

VIII. Summary of Part I and General Predictions 95
Specific Interaction Hypotheses 95
General Predictions 98

Part II: Analysis of the 1970-71 Planned Variations Data 100

Introduction 100
I. Initial Ability 107
The 70-71 Planned Variations Data: Results of Hypothesis Testing 108
Discussion 114

II. Previous Preschool Experience 118
The 70-71 Planned Variations Data: Results of Hypothesis Testing 119
Discussion 123

III. Sex 127
The 70-71 Planned Variations Data: Results of Hypothesis Testing 128
Discussion 130

IV. Age 134
The 70-71 Planned Variations Data: Results of Hypothesis Testing 135
Discussion 136

V. Socio-Economic Status 137
The 70-71 Planned Variations Data: Results of Hypothesis Testing 140
Discussion 141

VI. Ethnicity 145
The 70-71 Planned Variations Data: Results 146
of Hypothesis Testing 146
Discussion

VII. Response Style 151
The 70-71 Planned Variations Data: Results of Hypothesis Testing 152
Discussion 153

Part III: Discussion 158
General Predictions and Findings 158
Summary and Conclusions 162

Appendices

Appendix A: Correlations among Dependent and Independent Variables, 1969-70 Data 175

Appendix B: Program Effects of Hertzig-Birch Responses at Time of Pretest 177

Appendix C: Additional Notes on Sample and Methods 182

Appendix D: Additional Notes on the Analysis of the 1970-71 Planned Variation Data 184

Appendix E: Correlations among Dependent and Independent Variables, 1970-71 Data 193

Bibliography 195

Abstract

This report addresses the issue of whether different preschool programs have different cognitive effects on different types of children. Specifically, it focuses on three inter-related questions. First, what characteristics, or types of characteristics, of children interact most powerfully with characteristics of pre-school programs? Second, what are the patterns of such interactions? Third, and most broadly, how important are these interactions in explaining the cognitive outcomes of different programs? These questions are investigated through a selective review of relevant literature and through analysis of data generated in the first two years of the Head Start Planned Variations Study.

The report is in three parts. Part I proposes and justifies hypotheses concerning the interaction of particular preschool programs and particular characteristics of children. The hypotheses are based on the findings of previous comparative preschool evaluations and on analysis of data generated in the first year of the Planned Variations Study--1969-70. Part II reports the results of testing these hypotheses on another body of data--that generated in the second year of the Planned Variations Study--1970-71. Part III discusses patterns emerging from the two analyses.

The programs of preschooling examined are the eight involved in both the first and second year of the Head Start Planned Variations Study. They represent a wide spectrum of approaches to early education:

a. Academic Preschool Program: Englemann-Becker.
b. Behavior Analysis Model: Bushell.
c. Cognitive Model: Weikart.
d. Parent Educator Model: Gordon.
e. Tucson Early Education Model.
f. Responsive Model: Far West Laboratory.
g. Open Education Model: Educational Development Center.
h. Bank Street Approach.

All programs are examined separately in Part I, to generate hypotheses. In Part II, the patterns emerging from the first year of Planned Variations data--1969-1970--are used as a basis for grouping certain models together on a continuum running from "more-directive" to "less-directive". These categories are used to test certain hypotheses proposed in Part I, and are explained and justified in Part II.

The children's characteristics examined here are:

1. initial ability.
2. previous school experience.
3. sex.
4. age.
5. socio-economic status.
6. ethnicity.
7. response style: particular aspects of the ways in which children respond to the cognitive demands presented in the Stanford-Binet pre-test, as measured by the Hertzig-Birch method of scoring.

The outcome measures are two cognitive tests: the Preschool Inventory and the Stanford Binet IQ test.

The author concludes that because of the limitations of the two measures used to assess cognitive gains, and because of the many inconsistencies between patterns of achievement for the two tests and the two years of data, extreme caution is necessary in evaluating the educational significance of the patterns emerging from the analysis. Nevertheless the report suggests some tentative answers to the three questions which motivate the study.

First, across the two years of data, some children's characteristics interact more powerfully than others with characteristics of preschool models. Specifically, prior preschool experience and children's style of response to testing (as indicated by the Hertzig-Birch scoring of the Stanford-Binet) interact in a reasonably consistent way with the models--across the two years of data, although not always across the two cognitive measures used. Ethnicity and socioeconomic status, by contrast, do not show consistent and substantial interactions across the two sets of data. For three other children's characteristics, sex, age, and initial ability, the picture is mixed: interesting interactions of some magnitude emerge from the second year's analysis, but as these are unreplicated, they are treated cautiously. The author notes that the characteristics that show consistent interactions with the models--prior preschooling and response style--are the variables which relate most directly to the child's behavior and experience in the classroom, and which are most likely to change over time.

Second, in general, where interactions are strong or consistent across the two years of data they follow suggestive patterns. Specifically, for particular cognitive measures "more-directive" models seem to favor the achievement of children without prior preschool experience, those whose initial achievement on the Preschool Inventory is below the sample mean, and those whose response style is less mature. These patterns both parallel and qualify interactions reported by Bissell (70) and Bar-Yam (69). These researchers find children who might be described as "educationally-disadvantaged"

achieving more in highly directive environments than in less-directive ones. Bar-Yam's research relates to student ability-level, while Bissell's pertains to socio-economic status, but the patterns are similar.

Third, the Planned Variations analyses indicate that interactions may be quite important in understanding the cognitive outcomes of pre-school programs. None of the preschool programs is consistently more effective in raising the test scores of all types of children across both the two cognitive measures used (although one program does appear to be far more successful than others in boosting IQ). Furthermore, when particular models are grouped as "more-directive" and "less-directive" the main effect of model-group explains far less unique variance than do interactions of model-group with background characteristics. This suggests that monolithic prescriptions and monolithic categorization of children may be an unlikely avenue for educational advance. It also suggests the need to explore educational approaches that cater to the variety of children in every classroom, even classrooms in programs designed specifically for the poor.

INTRODUCTION

Evaluation of the outcomes of preschool programs raises at least three types of questions. First, the question of overall effectiveness: which program or type of program produces the greatest change on one or more measures, or which produces significant change on the measures believed to be most important? Second, do different programs have different patterns of outcomes--some producing cognitive change, for example, some leading to emotional growth, and others to progress in the areas of sensory development and motor control? Third, do the different programs produce maximum gains with different types of children, some working well in the south, and others in the north, some being more effective with very young children, or those of high ability, etc.?

These questions are of course not separate and distinct, since if outcomes of different programs actually are quite different, or if particular programs tend to be more effective with one type of child than another, then estimates of overall effectiveness will change when outcome measures or sample selection procedures change.

Out of the considerable body of research done on preschool programs, some fairly consistent patterns have emerged about overall effectiveness, and also about the range of effects we can and cannot hope for from preschool programs. The question of whether particular characteristics of children interact with program characteristics has been addressed in a number of studies (see, for example, Bissell, 70; Hodges, Spicker and McCandless, 67; Karnes, 69; and Kraft, Fuschillo

and Herzog, 68). For one child-characteristic, socio-economic status, a reasonably consistent pattern has emerged across a number of studies (Bissell, 70). For most other characteristics findings of different studies have been ambiguous or contradictory.

The fact that few consistent, well-documented interaction patterns have emerged to date may be in itself evidence on the question of whether first-order[1] interactions are significant in explaining the results of different programs. However, it seems at this point important to try to explore the interaction problem with another large sample of children and programs. The information generated by such an exploration could be important both for practical and for theoretical reasons: solid information on what types of program best promote cognitive growth for particular types of children can help administrators and parents choose programs for specific groups of children and can aid teachers in tailoring classroom practice to the needs of individual children. On the theoretical level, empirical data on the type of interaction that occurs between particular educational models and specific characteristics of children improve our understanding

---

[1]One reason for contradictory interaction patterns reported in different studies (several will be noted later; they are especially plentiful in the area of sex-by-model interactions) may be second-order interactions. Thus, for example, program A might show large gains for girls of high ability and boys of low ability, while program B favored girls of low ability and boys of high ability. If the sample was balanced by sex and ability, analysis would reveal no first-order interactions--neither sex-by-model nor ability-by-model. A second study, in which all subjects were of low ability, would show a strong sex-by-model interaction; a third study, using subjects of higher ability would show sex-by-model effects in exactly the opposite direction. This example is perhaps too tidy to be realistic, but it illustrates how second-order interactions might produce apparently contradictory patterns when only one child-variable is considered.

of the relation between classroom practice and children's learning. Finally, this type of information can affect the design of future evaluation studies, and the interpretation of those done in the past. Thus, an improved understanding of interactions can increase the precision of our estimates of preschool programs' overall effectiveness.

The present report focuses on the interaction question, attempting to answer, or begin to answer, three interrelated questions. First, what characteristics, or types of characteristics, of children interact most strongly with characteristics of programs? Second, what are the patterns of such interactions as do appear? Do they relate at all to the schemes and considerations which have been used in the past to classify preschool programs? Third, and most broadly, how important are interactions between characteristics of children and those of preschool programs? Can it be said that there is one "best" program for Head Start's target population, or rather that the program which most effectively promotes cognitive growth for one group of children is not the one which works best for another?

Design of the Study

These are very broad questions. The present report attempts to illuminate them by exploring the relation between a finite number of child characteristics, preschool programs, and outcome measures. Because the preschool literature sheds a rather uncertain light on several of the interactions under study, hypothesis generation has been seen as a major task of this report. And although certain preschool studies have proved to be invaluable resources, the approach to the problem has had to be in part empirical.

For this reason the report is in three parts. Part I uses the results reported in selected preschool studies, and the data generated in the first year (1969-70) of the Head Start Planned Variations Study[1] (PV) as a basis for proposing specific hypotheses concerning the interaction of preschool programs and child characteristics. Part II describes the results of testing these hypotheses on another body of data--that generated in the second year (1970-71) of the Planned Variations Study. Part III discusses patterns emerging from the two analyses.

---

[1]Planned Variations is a three-year study in which twelve different sponsors have been given funds and facilities to implement their model of preschooling in Head Start centers in several sites across the country. Data have been collected on children in these sponsored classrooms and in comparison classrooms (Head Start classes in which no sponsor is attempting to implement his model of preschooling). These data include both demographic information and pre- and post-scores on a variety of instruments.

For a description of the design and selected analysis of year 1 of Head Start Planned Variations see Implementation of Planned Variations in Head Start, SRI, 1971.

For a description of measures used in all the years of HSPV, see Walker, Bane and Bryk, The Quality of the Head Start Planned Variation Data, Huron Institute, 1972.

A. Models

The models of preschooling are the eight included in all three years of the PV study:

1. Academic Preschool Program: Engelmann-Becker (E-B)
2. Behavior Analysis Model: Don Bushell, University of Kansas
3. Cognitive Model: David Weikart
4. Parent Educator Model: Ira Gordon, University of Florida
5. Tucson Early Education Model: University of Arizona
6. Responsive Model: Far West Lab
7. Open Education Model: Education Development Center (EDC)
8. Bank Street Approach: Bank Street College of Education

The first two models listed are based, in quite different ways, on behaviorist assumptions and techniques. Bank Street, the last on the list, is the model which probably comes closest to the "child development" approach which has flourished for more than thirty years in middle-class nursery schools and--more recently--in many Head Start centers. The other five models are difficult to characterize briefly; all of them offer some structure through which teachers and aides may provide for children's cognitive development. The methods for accomplishing this goal are many, but it is probably fair to say that Gordon's program places major emphasis on contacts with parents, while Tucson, EDC, and Far West rely more heavily on a three-way encounter between child, teacher and materials. (The degree to which such encounters are planned in advance is not always clear from model descriptions and may vary considerably from classroom to classroom, as well as model to model.) The Weikart program is based on a Piagetian

PS 00860

framework of cognitive development, and emphasizes, among other things, the sequencing of learning experiences according to such a framework. Because any attempt to characterize eight sophisticated approaches to early education in a few sentences is bound to mislead, the reader is referred to the summaries in Maccoby and Zellner (70), and in the Head Start Planned Variations Study (71).

These models have been categorized in a variety of ways by other researchers. Stanford Research Institute (71) has used three categories: "Pre-academic", which includes Engelmann-Becker and Bushell; "Cogni[illegible]e Discovery", which includes Weikart, Gordon, Tucson and Far West; and "Discovery", which takes in Bank Street and EDC. David Weikart, who uses a matrix describing who (teacher, child, both, or neither) initiates learning experiences, groups some of them as follows (Weikart, 72):

> Programmed curricula (teacher initiates): Engelmann-Becker;
> Child-centered curricula (child initiates): Bank Street, Tucson, Far West;
> Open Framework (both initiate): Weikart.

He probably would add Bushell's model to the "programmed" category and EDC's to the "child-centered" group.

Rochelle Mayer, in "A Comparative Analysis of Preschool Curriculum Models" (1971) constructs a typology of models which depends on the importance a model places on each of three types of interaction: teacher-child, child-child, and child-material. Mayer examines four types of preschool model. The "verbal-didactic" (Bereiter-Engelmann) ranks highest on teacher-child interaction; the child development model places primary emphasis on child-child interactions; the "sensory-cognitive" or Montessori model stresses child-materials contacts more

than any other model examined. The fourth model she examines, Weikart's "verbal cognition" program, stresses all three types of interaction, concentrating on none to the exclusion of another. Since many combinations of emphasis are possible, this does not constitute an exhaustive list of model types.

Although no two researchers are in complete accord about model groupings, all agree that Bereiter-Engelmann and Bank Street, for example, represent quite different approaches. A concept commonly (though not universally) invoked in describing the differences between these programs is that of "structure".

The word "structure", as Mayer's careful analysis (1971) illustrates, has served an invaluable function in focusing attention on real differences and similarities between approaches. But, as Mayer points out, this use of the term "structure"--especially when models are contrasted as more or less structured--is misleading. Many models which are termed "less-structured" are, in another sense, highly structured: in the area of greatest program emphasis, teachers are carefully planning and sequencing children's experiences.

For this reason I prefer to use the less ambiguous term "directiveness" to describe the difference between the models of certain PV sponsors. This term refers to the extent to which teachers and other adults decide for the child how he will spend his time in school. There are, of course, no models which are totally "non-directive", since the way in which the classroom is set up almost always makes some activities more possible--and more attractive--than others, even in the absence of adult intervention. Beyond this, <u>all</u> teachers

establish and attempt to enforce some rules about children's conduct; nearly all teachers have some sort of schedule determining what activities or options are available at particular times. There are considerable differences between models concerning the extent to which sponsors expect adults to oversee children's choices directly and personally--to tell each child what to do each day. Thus, in models which are termed "less-directive", children make significant choices about how and with whom to spend time, for at least a part of each school day. In "more-directive" models, adults make many more of these decisions for children.

It would not be possible to place all PV models on a "directiveness" continuum (Gordon's Parent-Educator model would, for example, be hard to locate) but it does seem clear that Bushell and Engelmann-Becker are significantly more directive, at least in conception, than Bank Street, EDC, or Far West.

Much of what I have said about these educational approaches is theoretical--it is based on sponsor's model descriptions. How different are models in terms of what actually goes on in classrooms? We do not have the information needed to answer this question fully, but the Classroom Observation data, collected on PV and comparison classrooms, gives information on some aspects of classroom practice in the second year of the PV study, 1970-71. Specifically, this instrument describes the configuration of classroom activities at given intervals throughout the day, and the frequency of certain behaviors and types of interaction. We can learn from the observation data, for example, whether teachers in particular classrooms ask thought questions, how often

they inform children directly, and whether adults are more often with children individually, in small groups, or as a class. However, there is a good deal about the specific content of the encounters which the instrument cannot tell us.

Carol Lukas and Cindy Wohlleb, in their analysis of this data (72) find highly significant ($p < .001$) between-model differences on nearly all of the fifty-one observation variables. However, interpretation of these differences is often complicated by large differences between sites within models. It does seem clear from their analysis that on measures like "frequency of academic activity", "adults with children in academic activity", and "independent child activity", model differences are highly significant and pretty much in the direction which model descriptions would lead us to expect (See Lukas and Wohlleb, 72, for further description of the observation measure and exact figures on these and other variables). However, the many obstacles to implementation which they describe in their report would lead us to imagine that the subtler aspects of the model are not always in line with model descriptions. Since we have only a rough idea of how all forces come together to form actual classroom practice, we must be cautious in making assumptions about model differences.

In the analyses described in Part I of this report, programs have not been grouped in advance according to any categorization scheme. They have been examined separately for several reasons. First, and most important, the purpose of Part I is exploratory; it therefore seems counterproductive to limit analyses at the start through the use of a priori groupings. Second, given the documented problems in

implementing models on a national scale in the Planned Variations study (Lukas and Wohlleb, 72), both the models and their effects may be different enough to make the conventional labels inappropriate. Looking at outcomes of each program separately reduces the likelihood of obtaining statistical significance but may increase our understanding of the usefulness of particular types of categorizations in understanding the educational process.

Because the results of the analysis of the 69-70 data described in Part I suggest that the dimension of directiveness may be relevant to interactions with some variables, certain models are grouped together for particular analyses in Part II. These groupings are based primarily on sponsor model descriptions; they are described in Part II.

B. Characteristics of Children[1]

The child characteristics examined are as follows:

1. Initial ability
2. Previous preschool experience
3. Sex
4. Age
5. Socio-economic status (SES)
6. Ethnicity
7. Response style: particular aspects of the way in which

---

[1]These characteristics are described more fully in the chapters below.

children respond to the cognitive demands presented in the Stanford-Binet IQ test as measured by a procedure similar to the Hertzig-Birch scoring.

These characteristics were chosen because they represent a range of variables, both demographic and psychological, and because there was reason, either empirical or theoretical, to suppose that they might interact with characteristics of preschool programs.

C. Outcome Measures

The outcome measures used are two cognitive tests, the Preschool Inventory and the Stanford-Binet.[1] There are four reasons for this choice. First, these two measures have higher reliability than any

---

[1]The Preschool Inventory, developed in 1968 by Bettye Caldwell, is basically an achievement test, designed to measure knowledge in areas that are relatively independent of a child's particular background and experience. It tests the child's level of general information with items like "Where do you go when you are sick?" and "Color the triangle yellow". This test is very sensitive to maturation, with the greatest gains being made by the younger preschoolers in the PV sample Because scores for this test have not been normalized for age (as are the Binet scores) raw gains are harder to interpret. Thus, the child who gains six points on the PSI in eight months may actually have lost ground in relation to his age group (this would depend on his age); this is not true for the Binet.

The Stanford-Binet, although it does presuppose certain knowledge on the part of the child, is intended not to be an achievement test but rather a test of the child's learning ability. Many, although not all, of the items require the child to solve problems, both verbal and non-verbal: he is asked to duplicate a block bridge built by the experimenter, to identify missing or incongruent objects in a picture, to complete sentences by supplying opposites (e.g., "The day is light, the night is _____.")

The Preschool Inventory was administered to all children in the study, while the Stanford-Binet was given only to a randomly selected 50%. For this reason certain models are excluded from particular IQ analyses, due to small cell size.

others used in the PV study in 1969-70.[1] Second, they are more often used in preschool studies than the other tests used in the PV study, and hence are better known and most easily interpreted. Third, although the goals of the Planned Variations models vary greatly, all models include as an aim facilitating some kind of cognitive growth. Finally, previous research bearing on interactions between program and child characteristics has dealt largely in outcome measures of achievement and cognitive functioning. It seemed desirable to use this research in formulating hypotheses; therefore, it made sense to choose comparable measures.

Nevertheless, although these two tests are the best indices of children's cognitive development available to the Planned Variations Study, they are far from satisfactory if taken as complete measures of the cognitive effects of preschooling or of the degree to which PV sponsors have achieved their cognitive goals. This point has been made before (See Zimiles, 1970) but is repeated here as a caution in interpreting what follows. The limitations in the content of cognitive tests are striking. Although they can measure aspects of the knowledge and information which the child has acquired, they cannot measure the nature of his conceptual functioning--for example, the

---

[1]The test-retest reliability for the 1937 scale of the Stanford-Binet for children 2½ to 5½ ranged from .83 to .91, depending on the IQ of the child (the reliability is highest for children in the lowest IQ ranges). The reliability of the 1960 scale (used by PV) is probably higher, since only the most reliable items are included in the revised form (exact figures are not, however, available for the 1960 Revision).

The internal (KR-20) reliability of the PSI is about .90. For technical information on these two instruments, see Walker, Bane, and Bryk, The Quality of the Headstart Planned Variations Data, Huron Institute, 1972.

strategy a child uses for classifying and ordering events, or the ways he infers cause and effect. Furthermore, they say little about personality traits like initiative, flexibility, and perseverance, which many preschool planners consider a part of cognitive functioning, and which they intend their programs to foster.

In addition to these limitations of content, there are those of context. Many people--especially young children from low income families--do not function at their best during a test (Zigler and Butterfield, 68). This is true even when the tester is familiar and trusted; it is more true when he is a stranger (see Reisman, 62; and Labov, 69). Furthermore, models vary in the degree to which they prepare children for the didactic context of the test--complicating interpretation of data still further.

Neither the PSI nor the Stanford-Binet was designed to evaluate the success of a particular program in meeting specific goals. On the contrary, both tests are intended to be "curriculum-free"--appropriate for a wide variety of programs. They are intentionally insensitive to subtle differences between programs. Indeed, all PV sponsors have major cognitive goals which these instruments do not tap. Yet the subject matter of the tests is so broad that it is hard to be sure exactly what the tests do measure.

For all these reasons, the results reported in these pages must be interpreted with caution. When analysis shows one group of children outscoring another on the PSI in Model A, the temptation is to say that these children are "benefitting" more. But it is important to remember that we have evidence on only a very limited type of benefit. Often

the effect does not even extend to all cognitive tests, much less to other cognitive and non-cognitive goals of preschooling.

We use these cognitive measures for lack of other equally reliable instruments which measure other kinds of cognitive growth. If we keep their limitations firmly in mind, then they can give us certain kinds of information about effects of particular programs on different types of students. But if we assume, without any adequate evidence, that the results we are recording describe a pattern of total cognitive growth, we may well be misled.

PART I: HYPOTHESES

Literature Review and Report on the 1969-70 PV Data

Introduction to Part I

The strategy for generating hypotheses has two steps. The first is a selective review of the literature relating to interactions between the variables and the models under study. Two bodies of research seem particularly relevant. The first comprises the group of studies which, in evaluating the impact of particular preschool models, investigate the possibility of interaction between these models and one or more of the child characteristics on which this paper focuses. In a number of such studies, interactions are not explored explicitly, but the data presented suggests that a particular interaction has or has not occurred. In such cases the data is used to generate hypotheses.

A second group of studies, the aptitude-treatment interaction literature, examines the interaction of specific educational treatments with particular attributes of students. The attributes selected are, in general, more psychological than demographic--e.g., level of anxiety, compulsivity, general ability, specific abilities. In most of these studies the subjects are school and college students rather than preschoolers. This research is not used as extensively as the preschool studies, but where findings or hypotheses seem especially relevant they are summarized.

The second step in generating hypotheses is analysis and exploration of the data from the first year--1969-70--of the Head Start Planned Variations Study. In general, the strategy is as follows:

for each child variable, two-way[1] analyses of co-variance[2] are used to evaluate the importance both of the variable and of the interaction between the variable and the models in explaining the variance in post-test scores on the Stanford-Binet and PSI and gains[3] on these two tests.[4]

---

[1]The possibility of second-order interactions (see footnote on p. 2) suggests the advantages of using three-way analyses, rather than two-way. However, small cell sizes and an unbalanced design make most three-way analyses impractical.

[2]The Data-Text packaged program for unweighted-means analysis of covariance has been used. An unweighted means analysis was selected because the sample size for a particular program is unrelated to any real properties of the model. Since the number of children in each model is a matter of chance, there seems no reason to give greater weight to models which happen to have more children.

Covariates on these analyses include: age, sex, race, preschool experience, income, and family size. Pre-test score is included for post-test analyses but not gains analyses.

Additional information on the sample and analysis is given in Appendix C.

[3]The use of raw gain scores is currently in disfavor because of the problems deriving from unreliability of instruments--in particular regression-to-the-mean effect. The use of the post-test scores, adjusted by pre-test, is generally considered to be more satisfactory because it bypasses the regression problem. A strategy which includes the use of gain scores adjusted for covariates is used for reasons related to the design of the PV study. Because Planned Variations is not a true experimental design, with random assignment of sites to sponsors, certain variables are confounded with model (see Smith, 73). In consequence, in the preliminary analysis described in Part I, it seems necessary to adopt a conservative strategy: we use two types of analysis for each test and limit hypotheses to occasions where two strategies--or else effects on two instruments--show a measure of agreement.

Because in Part I results of gain score analyses appear similar to results of post-test analyses, the gain scores are not used at all in Part II. Techniques employed on the 1970-71 data are described in the introduction to Part II.

[4]For one child variable, age, a somewhat different approach is taken. Details are given in Section IV.

Because the purpose of Part I is to generate hypotheses, the pattern of interactions is explored even where the overall interaction does not reach acceptable levels of statistical significance.[1] In general, the results for the four different dependent variables are examined together, and where two or more analyses show substantial and compatible interactions between the variable and a particular model, a hypothesis is proposed. This criterion is based on the assumption that interactions which show up in only one analysis are less likely to reflect strong effects, and in consequence less likely to be replicated; it is, however, somewhat arbitrary.

The rest of Part I is divided into eight sections. Each of the first seven explores the main effect of one of the seven variables listed above and possible interactions of this variable with the eight models of preschooling. The format of these sections is similar: each begins with a brief description of the variable, followed by a summary of selected research which bears on this variable and its possible interactions with program type. A third subsection discusses results of analyses of the 1969-70 PV data as they relate to this variable. Finally, hypotheses relating to the variable are proposed. The hypotheses fall into two categories: strong and weak. Strong hypotheses are those grounded in earlier research and supported by results of the 1969-70 PV analysis. Weak hypotheses are of two types: some are generated empirically from the data with little or no grounding in previous research; others are based on the findings of

---

[1]Strictly speaking, "statistical significance" has little meaning in the context of the PV study, since the assumption of random assignment is violated (See Smith, 73). In this report, the term is used merely as a heuristic, to indicate the magnitude of effects.

earlier studies but are unsubstantiated by the first round of PV data.

The final section presents a summary list of hypotheses discussed in the preceding pages. This section re-examines the three questions raised in the introduction and proposes hypotheses relating to these questions.

I. INITIAL ABILITY

It seems reasonable to suppose that some educational approaches will produce their largest gains with children of high ability, while others will be most successful with the less able. For this reason it is of interest to compare the cognitive gains of children of high and low initial IQ in different models. There are, however, methodological difficulties in such comparisons: the regression-to-the-mean effect will artificially inflate the gains of children who score low on the initial test, while deflating those of high scoring children. This regression effect has two aspects, one of which is statistical and the other of which is "real".

Statistical regression is related to the reliability--or, more accurately, the unreliability--of the test used; for a perfectly reliable test (one for which pre-test score perfectly predicts post-test score) statistical regression poses no problem. However, for any test with a reliability of less than 1.0, random errors affect the precision of the score. Assuming no systematic bias exists, half these errors will lower children's scores, while the other half will raise them. If we look at the difference between pre- and post-test means for the whole population, these errors will balance out and the mean gain we observe will be the same as the real gain. If, however, we stratify the population according to initial test scores, dividing the children into, say, "high" and "low" groups, our low group will probably include more children whose scores have been artificially

depressed by testing errors, while our high group will include a disproportionate number of children whose scores have been inflated. Assuming that at post-test the errors are once again randomly distributed, the bottom group will appear to gain (even if no change at all has occurred) while the top group will appear to lose relative to them. The greater the unreliability of the test the stronger this effect will be.

There are several ways to deal with the problem of statistical regression. The first is to use one test to stratify and a second test to measure gains. This technique will reduce regression error to the extent that errors in the two instruments are uncorrelated. In dealing with the PV data, I have stratified children according to initial scores on the Stanford-Binet and then compared the gains of high and low IQ groups on another test--the PSI. We expect that errors on these two measures are as close to uncorrelated as is possible, since the tests are given on different days and by different testers. When this procedure is used, any unreliability in the initial IQ test score will tend to reduce differences between the groups, since it will lead some children whose "real" IQ scores are high to be classified as low, while others, who are really low, will be misclassified as high. This, however, will not basically distort results; it may lead to a more conservative estimate of the effect of initial IQ on gains.

The second way of dealing with the problem of statistical regression is to compare the standard deviation (SD's) of two groups, rather than the gains of high and low scoring children. The standard

deviation is here considered to be an index of the spread of scores; we would expect that if a program is systematically more effective in raising the scores of initially low-scoring children, then post-test scores would be closer to the mean than pre-test scores. Conversely, if the program is most effective with initially high-scoring children we would expect the range to increase. This increase or decrease in the spread of scores will be reflected in a parallel change in the standard deviation if the reliability of the test remains unchanged and if scores are normally distributed. If we have reason to believe these conditions are met, we may therefore make some (cautious) inferences by comparing pre- and post-test SD's.

These two tactics are helpful in separating statistical regression from any relation between "real" IQ and gains. However, the word regression is often used in preschool studies to describe a second phenomenon: the real tendency of children's measured IQ to rise during the first year of an effective treatment and then level off. Thus, a group of children who have made spectacular gains in the first preschool year (as, for example, in Weikart's comparative study) are expected to "regress" during the second year of the program.

This does not usually mean that the researchers believe the first-year post-test systematically inflated children's scores, but rather that the first year of preschool appears, generally, to provide a larger impetus to intellectual growth than subsequent school experiences, and that children who have made very large gains do not generally maintain the rate of growth relative to a national population of children of the same age. This phenomenon is real

enough, and worth bearing in mind, but in the interests of clarity it might be better identified by another name than regression. When I use the term regression in this report, I refer only to statistical regression.

Previous Research

Most studies of pre-school achievement do not present data which permit us to compare the relative gains of high and low scoring children and have complete confidence in our conclusions. When the researcher stratifies students according to initial scores and compares mean gains of the resulting groups, we find it difficult to separate regression effects from "real" differences. Comparing pre- and post-test SD's gives us somewhat more information, but in order to interpret the results of these comparisons we must assume that test reliability remains unchanged and that the shape of the distribution is nearly the same at pre-test and post-test. We cannot, of course, be certain this is true unless we have examined the data.

It would be justifiable to ignore data reported by other researchers and limit our inquiry to an examination of the PV data. However, it does seem desirable to compare PV results to those of previous research where this is possible. I therefore include results reported in other studies here where the likelihood of these results being misleading seems small.

If we look only at IQ data, the chances of the assumptions concerning distribution and reliability being met are good. According to the PV reliability study (Walker, Bane and Bryk, 1973), the reliability of this measure changes very little with pre-school experience.

A very slight increase in reliability from pre-test to post-test is to be expected due to the children's increased age. The shape of the distribution is harder to deal with: without seeing the original data we cannot be absolutely certain that the distribution remains unchanged. However, we do know that floor and ceiling effects, which are often observed on achievement tests, do not occur on the IQ due to the construction of the instrument. We also know that IQ scores do tend to be normally distributed. In the PV 1970-71 a practically distribution-free test (Nemenyi, 1969) has been used to test hypotheses of no difference in spread for each model, pre- and post-, on the PSI and the Stanford-Binet. Although both tests show some differences in distribution between pre- and post-test, results for the distribution-free comparison are the same as results for the simple comparison of SD's. These results suggest that the test is not terribly sensitive to small changes in the distribution such as are likely to occur on the Stanford-Binet. Nonetheless, we have obviously to be cautious in interpreting differences we observe.

The study by Kraft, Fuschillo and Herzog (68) of a two-year traditional nursery school program in Washington, D.C. illustrates the misleading potential of stratifying by IQ and then comparing gains of high and low scoring groups. An appendix to this study shows the lowest group (initial IQ less than 75) gaining 23.1 IQ points during the first year of the program; this is compared to a gain of 4.6 points for the highest group (initial IQ greater than 94). The author infers, not surprisingly, that the neediest children have benefited most. However, comparison of the variance in IQ scores

at the beginning and end of the first year reveals that, far from having decreased, the variance has increased slightly (not significantly).[1] I conclude from this fact that the reported differences between the gains of initially high and low-scoring groups are due in large part to regression.[2]

In Erickson's Kalamazoo (69) study, which compares the effects of a Bereiter-Engelmann program, a traditional program, and no program at all, the standard deviations of the IQ scores of both preschool groups are well above that of the control group at the end of the first preschool year (no pre-test scores given; random assignment of children among the three groups). The difference in variance between the Bereiter-Engelmann and control (no preschool) groups is significant at the .05 level; the variance reported for the traditional preschool group is not significantly different from those of Bereiter-Engelmann or control groups. For control children kindergarten is

---

[1]These results are puzzling, since the usually-high reliability of the Stanford-Binet would not lead us to expect a regression effect anywhere near large enough to explain these results. However, analysis of the 1970-71 PV data does suggest that the relation between fall and spring IQ score is different for younger and older children, with the correlation between pre- and post-scores being especially weak for young children. Children in the Kraft study are nearly a year younger at time of entry than the youngest children in the PV study. Therefore these odd results may simply reflect the great difficulties of obtaining a reliable test score for very young children.

[2]The reliability of the test may have changed between pre- and post-test but if it did it is likely to have increased (it is slightly easier to obtain a reliable score for children who are older and for those with prior preschool experience (See Walker, Bane and Bryk, 1973). An increase in reliability would be expected to decrease the SD, not increase it, so reliability changes probably do not explain reported results.

the first school experience; for the group entering the Bereiter-Engelmann kindergarten the variance in scores increases as does the mean IQ. For those attending the regular kindergarten the mean IQ and the IQ variance decreases somewhat. These results may suggest that when the first preschool year raises IQ scores substantially it is likely to increase the spread of children's scores.

David Weikart's (72) comparison of three preschool curricula--his own cognitive program, a Bereiter-Engelmann model, and a traditional program--shows dramatic increases both in mean IQ score and in IQ variance between the beginning and end of the first preschool year. It is interesting that the variance in post-test IQ scores of the traditional group is consistently below that of control groups (means and standard deviations are given for two control groups in Weikart, 70) while that of the cognitive group is consistently above that of control groups. These results suggest that Weikart's cognitive program tends to increase the spread in IQ scores while the traditional program--more than no program at all--acts to decrease it. However, I would interpret these data very cautiously for two reasons. First, the exceedingly small SD's reported at pre-test, and the increases reported for contrast as well as preschool groups, suggest that the changes may result in part from the initial selection process (only "functionally retarded" children). Second, the pre-test mean for all groups is so low--ranging from 73.0 to 84.4--that the whole group may be considered equivalent to low IQ groups in other studies.

Van de Riet's (72) evaluation of Sprigle's Learning-to-Learn program compares the IQ scores of four and five year olds before and

after participation in two preschool programs, one experimental and one traditional. For both groups of children in traditional programs (a nursery day care center and a Title III kindergarten) the standard deviation of scores increases markedly; in the case of the kindergarten children this pre-post difference is significant. Pre-post differences for the two experimental groups are smaller and insignificant; for the four year old group the SD actually decreases.

Hodges, Spicker, and McCandless (67), comparing gains of Appalachian five-year-olds enrolled in a structured experimental preschool, a public kindergarten, and no program at all, report no significant relation between gains and initial IQ for any group (standard deviations not given).

These data from preschool evaluations suggest that successful preschool programs, whatever their curricula, do not generally decrease the variance in IQ significantly, at least for first-year children. On the contrary, where a significant change in variance does occur it is generally an increase. In all likelihood this means that IQ gains are either uncorrelated with initial ability or that children of high ability are gaining more. The preschool data do not suggest that certain types of programs consistently increase the variance more than other programs. Although both Weikart (71) and Erickson (69) report higher post-test SD's for Bereiter-Engelmann than for traditional programs, in neither case are the differences significant. There is some suggestion, from David Weikart's comparative evaluation (Weikart, 71), that his cognitive program may increase the variance in IQ scores more than other pre-school programs.

Two sources might have led us to expect different patterns for traditional and Bereiter-Engelmann programs. The first is Bissell's re-analysis of the data collected by Karnes, DiLorenzo, and Weikart (Bissell, 70). Bissell finds an interaction between program and SES, with the higher SES children gaining more than lower SES children in "enrichment" programs and the lower gaining as much or more than the higher in "structured" programs (Bereiter-Engelmann and Karnes' Amelioration program). Since for nearly all programs the prescores she reports for low SES children are lower than those of the highest SES group, we might expect to find a similar pattern for initial IQ; in the terms of the present discussion this would lead to the expectation of a decrease in the variance of IQ scores for children in Bereiter-Engelmann programs and an increase for children in traditional programs; this expectation is not confirmed.

The aptitude-treatment-interaction (ATI) studies relating to general ability would lead to a similar expectation, although the relevance of these studies to PV models could be questioned. Bar-Yam's (69) review of that literature summarizes three studies (Wispe, 1951; Calvin, 1957; Ward, 1956) reporting an interaction between teaching style and student's ability. In all these studies students of low ability achieved more in "directive" than "permissive" classrooms. In general, the curriculum made less difference for the students of high ability, but when the measure was understanding rather than recall, the brighter students did somewhat better in the "permissive" environment.

One reason for questioning the pertinence of Bar-Yam's "directive"-"permissive" continuum to Planned Variation models is that in these ATI studies the class works as a unit in both the directive and non-directive treatments. In the directive method the teacher takes all of the responsibility for raising questions and presenting material, while in the non-directive method students take some of it. However, in the less-directive PV models most teaching is done either on a one-to-one basis or in very small groups, not in large student-directed classes. Research in the "discovery method" would be somewhat more relevant, but investigators studying "discovery" classrooms have not generally used ability to categorize students.

For all these reasons plus some others--the age of the students, the outcomes sought--the ATI literature has limited application to the PV study. However, these studies do illustrate the direction of thinking among researchers who have considered possible interactions between classroom practice, intellectual ability, and achievement. One purpose in examining the 1969-70 data is to see to what degree and in what ways the hypotheses and distinctions of this literature are relevant to PV models.

1969-70 PV Data

Comparison of the standard deviations in pre- and post-IQ scores reveals that, for the PV sample as a whole, and for each individual model except Weikart, the SD's decrease between fall and spring testing. For the sample of all children in PV classrooms, this drop is

significant ($p < .05$), but for most individual models it is not, due to the smaller N's. The exception is Gordon, where the decrease in variance is significant ($p < .05$). The variance in IQ scores also drops for comparison children (children in unsponsored Head Start classrooms) but the drop is considerably smaller and insignificant. Were this not the case we might conclude that the pre-post variance changes result merely from an increase of test reliability. However, the difference between PV and comparison samples indicate that this explanation is inadequate.

Table Ia gives means and SD's pre- and post- for all models. For the Stanford-Binet these figures are given first for all children in the model and then for those without previous pre-schooling. The pre-post SD's are compared both ways on the theory that the drop in SD's might merely reflect first-year children catching up with those having previous preschool experience. However, this appears not to be the case: the decrease is no less for the group of children without previous preschooling than for the sample as a whole.

The PSI also shows SD's decreasing across models. These results suggest that for both tests children who initially score low gain more than those who score high. They are surprising inasmuch as programs studied by other researchers have shown a tendency to increase rather than decrease the variance of scores of first-year children.

For further analysis of the PSI data, children have been divided into two groups on the basis of their fall IQ scores, the "high" group consisting of children initially scoring above the sample mean (92.5 points), and the "low" group of children scoring below it. The

Part I

Table Ia*

Stanford-Binet and Preschool Inventory
Means and Standard Deviations on Pre- and Post-tests
for Children in Eight Models
1969-70 PV Data

Stanford-Binet All Children	Far West	Tucson	Bank St.	E-B	Bushell	Weikart	Gordon	EDC
pre $\bar{x}$	90.063	95.625	90.729	91.813	96.000	98.256	88.029	91.277
SD	12.709	13.070	9.398	10.323	16.234	11.727	15.143	15.001
post $\bar{x}$	98.500	100.563	96.438	94.563	100.786	101.953	94.200	93.319
SD	8.618	11.947	8.243	9.500	10.864	13.413	11.324	13.371
N	16	48	48	64	14	43	35	47

*Sample: all children with valid scores on both Stanford-Binet and PSI, fall and spring.

Table Ia (continued)

Stanford-Binet Children without Prior Preschooling	Far West	Tucson	Bank St.	E-B	Bushell	Weikart	Gordon	EDC
pre $\bar{x}$	90.063	93.306	90.804	93.222		98.741	84.269	88.543
SD	12.709	13.957	9.598	10.849		10.790	14.115	12.342
post $\bar{x}$	98.500	98.861	96.609	96.333		102.815	92.923	93.152
SD	8.618	11.733	8.381	8.958		12.782	11.267	11.264
N	16	36	46	9		27	26	33
PSI Children without Prior Preschooling*								
pre $\bar{x}$	35.519	34.727	36.253	37.880	36.625	41.767	34.887	36.058
SD	11.091	11.471	12.280	11.308	10.474	11.238	11.133	11.316
post $\bar{x}$	46.222	45.039	45.759	52.720	53.375	53.810	51.849	50.115
SD	8.631	10.398	10.322	9.158	7.425	8.360	8.179	7.883
N	27	77	166	25	8	116	53	52

*SD's not given for whole sample, since a slight ceiling effect on the PSI for older children with prior preschooling would be expected to depress post-test SD's.

unweighted means analyses of covariance reveal significant main effects for initial IQ. As expected, these effects favor low IQ children for PSI gains ($p < .05$), the difference between the adjusted mean gains of high and low scoring children being 2.702 points. For the covaried post score the main effect ($p < .001$) favors high IQ children, with the difference in adjusted post-scores being 3.998 points. Evidently, low IQ children start lower, gain more, but still end up lower.

The model-by-IQ category interaction is significant ($p < .05$) for the analysis of PSI gains but not for the PSI post-test analysis. Interaction effects are computed after the main effects of model and IQ category are taken out; they refer to the within-program effect of the IQ category on adjusted (covaried) scores or gains. Thus, for the Tucson program the difference between the adjusted PSI post-test scores of high and low IQ groups is 8.068 points. Of this difference, 3.998 points is accounted for by the main effect of IQ. This leaves a 4.070 point difference to be explained by an interaction of IQ category and model; this interaction effect is expressed (see upper-right-hand cell of Table Ib) as ±2.035.

Table Ib below gives the magnitude and direction of all interaction effects greater than ±1.0 points for both analyses. This cut-off is chosen arbitrarily, as a heuristic. ±1.0 points expresses a difference of 2.0 points between adjusted post-test scores--.20 SD's on the Stanford-Binet, and .17 SD's on the PSI.

The 69-70 data suggests that across programs there is a tendency for children of low initial IQ to gain more than those of high initial IQ both on the PSI and on the Stanford-Binet. It suggests that this

Part I

Table Ib*

Analysis of Covariance Interaction of Initial IQ Category with Model:
Adjusted Interaction Effects Greater than 1.0 Points**
1969-70 PV Data

	Positive for low IQ children (initial IQ < 93 points)	Positive for high IQ children (initial IQ > 92 points)
PSI Post-test	Weikart (± 1.136) Gordon (± 2.047)	Tucson (± 2.035) Bank St. (± 1.122)
PSI Gain	Weikart (± 2.605) Gordon (± 4.643)	Bushell (± 1.042) Bank St. (± 1.568) Tucson (± 2.952) EDC (± 1.41)

*Sample: all children with valid scores on PSI, fall and spring.

**This means that with the main effect of previous preschooling taken out the difference between adjusted means is greater than 2.0 points. This amounts to one-fifth of a standard deviation on the PSI and one-sixth of a standard deviation on the Stanford-Binet.

Covariates include age, sex, race, preschool experience, income, and family size; pre-test score is a covariate in the post-test analysis. All models are included in these two analyses.

tendency is strongest and most consistent in Gordon sites. Beyond this, it offers some support to the ATI hypothesis that less-directive programs are more favorable to the achievement of high ability students than to that of low ability students: the programs showing consistent effects of some magnitude favoring high IQ children for the PSI are all on the less-directive end of the continuum--see Table Ib, right-hand side (Bushell does show up in the lower-right-hand cell, but not in the upper right).

The results for the Weikart program are contradictory and somewhat baffling. The IQ data suggests that this may be the only PV model to increase the variance in IQ scores between fall and spring testing. This fits with Weikart's own data (Weikart, 72) which shows the variance in IQ scores for children in his cognitive program rising significantly and dramatically during the first preschool year. The PSI data, however, shows an effect favoring the achievement of low IQ children--the exact opposite of what the IQ data would lead us to expect.

If this contradictory pattern were to be repeated in the 1970-71 data we would be forced to see it as resulting from differences between the two tests, saying in effect that this model increases the spread of IQ scores but decreases the spread of achievement. But until such an interpretation receives support from another round of data, I would give greater weight to the IQ data because it follows the pattern of Weikart's earlier study.

Hypotheses

Strong

1. Less-directive models will show interactions favoring children of high initial IQ on the PSI post-test.

   The ATI literature suggests the hypothesis that in less-directive ("permissive") environments, children of low ability will be at a disadvantage. This hypothesis is supported by the 69-70 data.

2. Weikart programs will show a stronger tendency than other programs to increase the variance of IQ scores.

   This hypothesis is suggested by Weikart's comparative evaluation data (Weikart, 72) and supported by the 69-70 data.

Weak--Based on 69-70 PV Data

3. In Gordon programs within-model interaction effects on the PSI will favor children of low initial IQ. The variance of IQ scores will tend to decrease in this model.

II. PREVIOUS PRESCHOOL EXPERIENCE

Both in 1969-70 and in 1970-71, the Planned Variations sample includes a number of children with some prior preschool experience. Although the PV study does have information on the duration of this preschooling, and whether or not it was Head Start, the study has no data concerning the character of this experience--whether the program was highly-directive or not, whether the orientation was academic, emotional, or simply custodial. It is therefore unfortunately necessary to lump together all prior preschooling, even though children's experiences undoubtedly differed in important ways.

In the 1969-70 PV sample, children with previous preschooling score higher on the Stanford-Binet pre-test than children without it. For the entire 69-70 sample, the magnitude of this difference is 5.4 points. As we might expect, however, the raw gains of the children without preschool experience are larger; so that post-test differences between the two groups are considerably smaller.

The relatively small IQ gains made by PV children in their second year of preschooling are not particularly surprising: the data from most other evaluations of two-year programs follow a similar pattern (see, for example, Gray and Klaus, 1968; Kraft, Fuschillo, Herzog, 1968; Beller, 1969; Erickson, 1969; Karnes, 1969; Weikart, 1971; Van de Riet, 1972).[1] Even in Beller's study, which is remarkable in

---

[1]An interesting exception to this rule is the control group in Van de Riet's study (72). This group made no IQ gains in a year of a traditional day care program, but made significant gains in a Title III kindergarten the following year. The other control group, without previous preschooling, showed no IQ gains in the kindergarten.

reporting significant IQ differences between children with one and two years of preschooling which are maintained as far as third grade, kindergarten IQ gains of children with nursery school experience are very small.

This does not mean, however, that the effect of a second preschool year on children's test scores is trivial. On the contrary, as Marshall Smith (73) has shown in his report on the 1970-71 PV data, even children with prior preschooling show gains two to three times as great as those we would expect to observe for a comparable group of children not enrolled in preschool.[1] This is true both on the PSI and on the Stanford-Binet.

From what we know about the emphases of different models, it seems reasonable to guess that some programs may be especially effective in boosting test scores of children who are entering preschool for the first time. Other programs--perhaps those which place a high value on fostering children's initiative--might be somewhat more successful with second year children. I have therefore asked two questions concerning the pattern of second year gains. First, are some models more successful than others in producing cognitive gains for children with an earlier preschool experience? Second, do some types (or categories) of children gain more than average from a second preschool year?

---

[1] This finding is explained in M.S. Smith, Some Short-term Effects of Project Head Start: A Preliminary Report on the Second Year of Planned Variation, 1970-71, Huron Institute, January 1973.

Previous Research

The effect of curriculum on second year gains. The design of the Kalamazoo study (Erickson, 1969) makes it possible to study the relative importance of first-year curricula, second-year curricula, and the match between the two. Children in this study were randomly assigned to three preschool treatments: Bereiter-Engelmann, traditional, or none (a control group). After one year of preschool, half the children in each group were assigned to a Bereiter-Engelmann kindergarten and the other half to a traditional kindergarten. The IQ scores at the end of preschool and kindergarten are given below for each combination of school experiences.

Table IIa shows a strong main effect of first year curriculum on second year IQ post-score; this effect favors children from the Bereiter-Engelmann preschool. The main effect of second-year curriculum is insignificant. However, the interaction of first and second year curricula, taken in conjunction with first year curriculum, is significant ($p < .05$) according to Erickson's regression analysis. Table IIa indicates that children in each preschool group benefited more from a kindergarten experience which was unlike their preschool than from one that was like it.[1]

Who gains most from a second preschool year? The IQ data reported by Kraft, Fuschillo, and Herzog (68) for

---

[1]Two other studies (Karnes, 1969; Weikart, 1972) compare the IQ gains of children in several types of preschool program over two years. However, because children were in the same program both years in these two evaluations it is impossible to separate the effects of first and second year curricula on second year IQ gains.

Part I

Table IIa

Stanford-Binet Scores for Children in Two Types of Preschool and Kindergarten Program (Adapted from Erickson, 1969)

	Preschool Curriculum		
	Bereiter-Engelmann	Traditional	Control
IQ at end of preschool	136	138	
$\overline{X}$	108.1	105.7	
S.D.	17.9	16.7	
IQ at end of traditional kindergarten	30	28	29
$\overline{X}$	111.7	100.6	91.5
S.D.	13.89	13.52	9.33
IQ at end of Bereiter-Engelmann kindergarten	30	28	29
$\overline{X}$	108.7	103.2	104.9
S.D.	13.36	13.07	16.53

children attending a two year traditional nursery school program in Washington, D. C. indicates that different children may benefit during the first and second years of preschool. For their sample, the probability of a child's making "significant" second year IQ gains (they define gains of five points or more as significant) was nearly twice as great if he had not made such gains the first year as if he had.[1]

The Kraft data also suggest that, at least in a traditional nursery school program, children of low initial IQ (where "initial" refers to IQ at the beginning of the first year of preschool) may make greater IQ gains during the second preschool year than children of high initial IQ. Kraft et al. report second year gains averaging 7.6 points for the group with initial IQ's of less than 75 and an average loss of .2 points for children whose initial score exceeds 94 points. These differences cannot be attributed simply to regression error, since the initial IQ on which children are classified is not one of the two IQ scores used in computing second year gains.[3]

---

[1]

		YR 2 Gains: Yes	YR 2 Gains: No
YR 1 Gains	Yes	7	16
	No	7	5

Adapted from table on p. 71 in Kraft, Fuschillo and Herzog.

[2]Page 76.

[3]Children were tested the summer preceding the first preschool year, during May and June of the first year, and during May and June of the second preschool year. The first test was used to classify them according to initial IQ; the second and third were used to calculate second year IQ gains.

The data reported in Erickson's Kalamazoo study suggests a similar pattern. Although gains are not reported in terms of initial IQ strata, inferences about what group has gained most from the second preschool year may be made by comparing the variances in IQ scores at the ends of the first and second years. If children who scored high on the earlier test have gained more than those initially scoring low, the variance would increase. If, on the other hand, the children who scored lower on the first test have gained more, the variance would decrease. For both preschool groups, the variances in IQ scores reported for the end of the preschool year are significantly higher than those reported for the end of the kindergarten year. This suggests that in both types of kindergarten, second year children of low IQ made greater gains than those of higher initial IQ.

One might guess from these two studies that children of high ability adjust more quickly to preschool than those of low ability and in consequence make somewhat greater gains during the first preschool year (as noted in the section on initial IQ, most studies show SD's increasing between fall and spring of the first preschool year). It is possible that because of their larger first year gains these high ability children are less likely to make second year gains; the children of low ability, on the other hand, may have taken longer to adjust to preschool and make large gains only in their second year.

Karnes, like Kraft, reports IQ gains by initial IQ strata as well as by program. In the Bereiter-Engelmann kindergarten, children of low and average initial IQ ("initial" referring, again, to the score

computed at the beginning of the first preschool year) gain considerably more during their second preschool year than children scoring relatively high at the beginning of preschool.[1] The significance of these differences can not be evaluated because SD's are not reported. No consistent differences between strata exist for the four groups attending public kindergarten.

Van de Riet's evaluation of Sprigle's "Learning to Learn" Program (Van de Riet, 1972) suggests that a second preschool year may actually have a long term effect on the variance in IQ scores. Van de Riet reports that although the mean IQ score of the children with one and two years in the experimental preschool program does not differ significantly by the end of grade 2, the standard deviation of the scores of the group with two years of preschooling is considerably (and significantly) smaller than that of the group with only one year of preschooling.

Kraft et al. suggest the possibility of a second-order interaction between socio-economic status, program and previous preschool experience. Significance levels are not given, and the numbers are small, but, again, the data is suggestive; Table IIb below, adapted from their presentation, shows the proportion of children at each SES level making gains of five or more points in the first and second year of the traditional nursery school program. The categories are not mutually

---

[1]Again, these effects cannot be attributed to statistical regression, since the initial IQ score used in stratifying the sample is not one of the two used in computing second year gains.

Table IIb*

Proportion of Children in Three SES Categories Making IQ Gains of Five Points or More (Greater Than One-half a Standard Deviation) during the First and Second Preschool Year.

		SES Level		
		Adequate (n = 8)	Borderline (n = 7)	Poverty (n = 19)
Proportion of children making gains of five points or more	Yr. 1	.75	.72	.58
	Yr. 2	.25	.42	.53

*Adapted from Kraft et al. (1968) table on page 76; children classified in accordance with income level categories devised by the Social Security Administration.

exclusive: some children gain both years and others make no gains at all.

Table IIb suggests a positive correlation between SES and IQ gains for the first year children. This is in line with Bissell's (70) finding that for less-structured programs SES is positively correlated with gains. However, the second year data reverses this pattern, and it appears that for the second year SES may be negatively correlated with gains, even in this less-structured preschool.

In summary, the data collected in previous preschool evaluation studies suggest the following hypotheses concerning the cognitive effects of a second year of preschooling and the interaction of previous preschool experience and program type:

1. In a given preschool program, the IQ gains of children with previous preschool experience are less than those of children without such experience (all studies).

2. Children with a particular type of preschool experience make greater second year IQ gains in a program which is quite different from their first experience than in one which is similar to it (Erickson).

3. Children whose first preschool experience is in a highly structured program are more likely to maintain or increase IQ gains during the second preschool year than children whose first year program is less structured (Erickson).

4. Children making large IQ gains in their first preschool year are less likely to make substantial gains during the second year than children making negligible gains the first year (Kraft, Fuschillo, and Herzog).

5. Children of low initial IQ are more likely than children of high initial IQ to make large gains during the second year of preschooling (Kraft, Fuschillo, and Herzog; Karnes; Erickson).

6. Children of low SES are more likely to make second year IQ gains than those of high SES (Kraft, Fuschillo, and Herzog).

1969-70 Data

Because specific information on the character of each child's earlier preschool experiences is unavailable, many of the hypotheses listed above can not be tested on the 1969-70 data. These data have therefore been used mainly to test hypothesis 1 and to look for interactions between particular models and preschool experience.

To evaluate the significance of interaction effects, two-way analyses of covariance were performed on PSI post-test scores, PSI gains, IQ post-test, and IQ gains.[1] Far West was excluded from all analyses, and Bank Street and Bushell from IQ analyses, because of the small number of children with preschool experience enrolled in these models. All four analyses show main effects favoring children without previous preschooling. These effects are significant only for the two analyses of gains ($p < .02$).[2] The magnitude of the effect is substantial for IQ gain, the unweighted adjusted gain for children with preschool experience being 1.1 points, as contrasted with 5.4 points for those without preschooling (a difference of about one-third of a standard deviation). The difference between the two groups for PSI gain is smaller, amounting to only 2.0 points (about one-fifth of a standard deviation).

---

[1] See Introduction, page 16 for covariates and general methods.

[2] In other words, children with prior preschooling gain less, on the average, than those without, but the difference between the two groups becomes insignificant when one controls for initial score. Apparently children with prior preschooling gain less only to the extent that they start higher.

The interaction of preschool experience and model is significant above the .05 level only in the PSI gain analysis, but approaches significance for the other three analyses. Table IIc shows the size and direction of interaction effects for the four analyses, where the magnitude of these effects is greater than 1.0 points. It is important to remember that programs listed as showing effects positive for children with previous preschooling (left-hand side of Table IIc) do not necessarily show larger adjusted gains for these children, since the main effect of previous preschool experience has been taken out. Thus, children with preschool experience in Engelmann-Becker sites actually gain less on the Stanford-Binet than those without such experience, but because the difference between the two groups is 1.3 points, as contrasted with the 4.2 points for the entire PV sample, it is said to show an interaction favoring those with preschooling.

On the basis of Table IIc we might hypothesize interaction effects favoring children with previous preschooling in Tucson and Bank Street, and favoring children without preschool experience in Gordon and Engelmann-Becker.[1] Weikart's scheme for classifying programs (Weikart, 72) might be useful in interpreting these observations. It seems reasonable to suppose that in the programs which require a good deal

---

[1]Ordinarily, models which show contradictory patterns on the two tests are not included in hypotheses. I make an exception in this case for two reasons. First, as explained on the previous page, the IQ gains interaction favoring those with previous preschooling is somewhat misleading--it actually reflects not a tendency of the model to benefit these children substantially, but rather a failure to produce average IQ gains for children without preschool experience. Second, the effects on the PSI favoring those without preschool experience are more substantial than any others on the table.

Part I

Table IIc*

Interaction of Prior Preschool Experience with Model
Analysis of Covariance: Adjusted Effects Greater Than 1.0 Points**
1969-70 PV Data

	Positive for children with previous preschooling		Positive for children without previous preschooling	
PSI Post	Tucson Bank St.	(±1.297) (±1.498)	Engelmann-Becker Gordon	(±1.429) (±1.816)
PSI Gain	Bank St.	(±2.149)	Engelmann-Becker Gordon	(±3.268) (±1.375)
IQ Post (Bank Street & Bushell excluded)	Tucson	(±2.907)	Gordon	(±1.207)
IQ Gain (Bank Street & Bushell excluded)	Tucson Engelmann-Becker Weikart	(±1.929) (±1.464) (±1.695)	Gordon EDC	(±2.669) (±2.419)

*PSI analyses include all models except Far West. IQ analyses include Tucson, Engelmann-Becker, Weikart, Gordon and EDC. The three other models include too few children with valid pre- and post-IQ scores who have prior preschool experience.

**Covariates include age, sex, race, income, family size, and, for the post-test analyses, pre-test score. See p. 32 for explanation of adjusted effects.

of initiative on the part of the child--"child-centered" programs, in Weikart's terms--previous preschooling might be an advantage, while in programs where the primary responsibility is laid on the teaching adults, such experience might be less useful. Weikart places both Bank Street and Tucson in the "child-initiates" category, and Engelmann-Becker in the "adult-initiates" group. Although Gordon differs importantly from all these models, its primary emphasis seems to be on the role of adults in children's learning. In this it seems to resemble those models falling into the "adult-initiates" group.

In effect this may mean that in "less-directive" or child-centered programs certain kinds of preparation facilitate cognitive gains, while in "more-directive" models preparation is less important--perhaps even a disadvantage. This result parallels the suggestion made in the previous section that less-directive models favor the achievement of children of high initial IQ. Taken together they indicate that children with certain kinds of educational advantages--either prior preschooling or higher initial IQ--may be more likely to make effective choices in a preschool environment.

There is, however, another possible interpretation, and it is worth bearing in mind. It may be that the choices made by the "more prepared" children are simply more likely to lead to the types of learning measured by the PSI and the Stanford-Binet. It is quite possible that the first-year children--and/or those with low initial IQ scores--are learning cognitive skills of equal or greater importance, but ones which go unmeasured. They may be learning how to find something to do, how to scan a room to learn what options it offers, how

to distinguish an adult who is free to give help from one who is busy or preoccupied, and how to tune out some, but not all, of the noises of a busy room. These skills, and a hundred others like them, are not measured by the PSI and Stanford-Binet so we don't know who is learning them, or in which models. But it certainly seems possible that first-year children in less-directive classrooms spend much time polishing these skills; equally, it makes sense that children who possess them may, in such an environment, make greater gains on cognitive tests.

Hypotheses

Strong

1. The cognitive gains of children having previous preschool experience will be smaller than those of children having no such experience.

   This finding is common to nearly all preschool evaluations (see p. 36). It is supported by the 69-70 PV data.

Weak--Based on Literature

2. Children with a particular type of preschool experience will make greater second-year IQ gains in a program which is different from their first experience than in one which is similar to it.

   This hypothesis is suggested by Erickson's IQ data.

3. Children whose first preschool experience was in a highly structured program will be more likely to maintain or increase IQ gains during the second preschool year than children whose first-year program was less structured.

   This hypothesis is suggested by Erickson's IQ data.

4. Children making large IQ gains in their first preschool year will make smaller gains during the second year than children making negligible gains the first year.

   This hypothesis is suggested by the findings of Kraft, Fuschillo, and Herzog.

5. Children of low initial IQ will make greater gains during the second year of preschooling than children of high initial IQ.

   This hypothesis is suggested by the data of Kraft, Fuschillo, and Herzog, by the Karnes data on IQ gains in the Bereiter-Engelmann kindergarten, and by the IQ variances reported in Erickson's study.

6. Children of low SES will make greater second year IQ gains than those of high SES.

   This hypothesis is suggested by the data of Kraft, Fuschillo, and Herzog.

Weak--Based on 1969-70 Data

7. In Tucson and Bank Street programs, interaction effects on both cognitive measures will favor children with previous preschooling; in Gordon programs, interaction effects will favor children without such experience. In Engelmann-Becker programs interaction effects on the PSI will favor those without previous preschool experience.

   This hypothesis is based on the 1969-70 PV data. In a more generalized form it might be stated "child-centered" or "less directive" programs will show interaction effects favoring second-year children while programs where the initiative lies primarily with adults will show interaction effects favoring first-year children.

III. SEX

Differences both in the maturation rate of preschool girls and boys and in culturally-based expectations for their behavior raise the possibility of sex-by-model interactions. However, contradictions in the research on intellectual differences between boys and girls in different preschool programs create formidable obstacles to the generation of specific hypotheses about the form of such an interaction.

Previous Research

The data on how (and whether) different types of preschool programs affect boys and girls differently is inconclusive. As a rule, boys score slightly lower than girls on cognitive pre-tests; on the Stanford-Binet mean differences usually amount to between one and four points. We might therefore expect boys to gain slightly more than girls, due to regression effects. If researchers control for pretest score the regression effect poses no problem, but unfortunately few studies do this. Instead they contrast raw gains for the two groups. In consequence, when patterns tend to reduce initial differences between the scores of girls and boys we cannot usually draw any conclusions.

Research on traditional nursery school programs shows no strong pattern of greater gains by either boys or girls. However, in the two studies I reviewed which show significant sex effects which cannot be attributed to regression, differences favor girls. Hodges, Spicker and McCandless (67) in their work with five-year old Appalachian

children report insignificant differences favoring boys for all groups (public kindergarten, experimental preschool, and controls). Kraft et al. (68) also report that boys gain more than girls in both years of their Howard University nursery school study. However, both these studies use raw gain scores, so it is difficult to tell how much of the observed effects are due to regression. Erickson (69) finds no difference between the gains of girls and boys in the traditional program he examines.

Bissell, in an analysis which controls for pretest score, SES, and ethnicity, reports effects favoring girls in the "permissive-enrichment" programs she examines. These differences are significant ($p < .05$) in one program and insignificant in the other two. Smith (68) also reports large and highly significant differences favoring girls at the end of a full-year pre-kindergarten program in Trenton (IQ differences between boys and girls not yet in preschool are insignificant). Neither Bissell's results nor Smith's can be attributed to regression.

Results reported for Bereiter-Engelmann programs are somewhat more consistent, although they do not suggest strong effects: Erickson (69) and Bissell (70) report small differences favoring boys in Bereiter-Engelmann programs. These differences are insignificant for Bissell's sample; significance levels are not given for Erickson's.

These results suggest three things. First, when sex differences in achievement are found in traditional programs--differences which cannot be attributed to regression or poor methodology--they are likely to favor girls. Second, the magnitude and significance

of these effects will vary considerably from school to school, or site to site. And third, sex differences in Bereiter-Engelmann programs will usually be small.

These patterns seem to make sense. Girls apparently enter preschool slightly more prepared (one evidence of this is their slightly higher pretest scores on cognitive tests). As I have suggested in previous sections, more prepared children may be at more of an advantage in less-directive models than more-directive models. However, these sex differences are not large. It seems entirely likely that in programs where guidelines for teacher-child interaction are flexible (this would describe most traditional programs), teacher expectations concerning sex differences and sex-appropriate behavior will influence the pattern of sex effects as much or more than real differences between girls and boys. Hence we should not be surprised to see the magnitude of sex effects varying considerably from subculture to subculture, region to region, and even classroom to classroom.

1969-70 Data

The analyses of covariance using the four dependent variables PSI and IQ post-score and gains yield remarkably consistent results: the main effect of sex is insignificant in all analyses, as is the interaction of sex and model. For the PSI post-score, all adjusted interaction effects are less than 1.0 point. Interaction effects for IQ gains and covaried post-test score are somewhat greater than those observed for the PSI, although still insignificant overall. Consistent effects favoring boys are seen in Far West and Tucson models; the

only model to show a consistent effect favoring girls is Bushell (see Table IIIa).

These data support the general impressions created by earlier studies: first, strong, consistent sex-by-model interactions are not to be expected in analyses of preschool outcomes, and second, the main effect of sex on cognitive outcomes is very small.

The 1969-70 PV data suggest that when testers and children are drawn from a variety of regions and subcultures, differences in response style, as defined by Hertzig-Birch codes, are almost totally uncorrelated with sex;[1] this may, of course, be untrue for particular classrooms, sites, or testers. Similarly, the overall correlations of sex with pre- and post-test scores on the cognitive measure is low, ranging from .075 to .165, although in particular sites it goes higher.

I think these data suggest that where main effects of sex or sex-by-model interactions seem strong they may have been created by particular local attitudes or teaching practices rather than by innate characteristics of children or models.[2] Beyond this, they raise the

---

[1]See Appendix A for exact correlations. Other researchers have looked for sex differences in the test-taking behavior of pre-schoolers with mixed results--Crandall and Rabson, in looking at whether children chose to return to tasks they had completed successfully or those at which they had failed, found no sex differences among three to five year olds (although sex differences were strong and significant among children aged six to nine). Moriarty, on the other hand, found significant sex differences in the speed at which preschoolers oriented to the Stanford-Binet testing situation, and in behavioral responses to the more difficult tasks.

[2]It is interesting to note that two of the three models demonstrating consistent sex-by-model interactions (Far West and Bushell) are one-site models in 1969-70.

Part I

Table IIIa*

Interaction of Sex with Model
Analysis of Covariance: Adjusted Effects Greater Than 1.0**
1969-70 PV Data

	Positive for Boys	Positive for Girls
PSI Post	---	---
PSI Gain	Weikart (±1.050)	Bank St. (±1.236) Bushell (±1.375)
IQ Post	Far West (±3.533) Tucson (±1.166)	Bushell (±2.657) EDC (±1.536)
IQ Gain	Far West (±2.809) Tucson (±1.542) Gordon (±1.622)	Bushell (±3.414) Weikart (±1.285)

*All models included in all analyses.

**This means that with the main effect of sex taken out, the difference between adjusted means is greater than 2.0 points. This amounts to one-fifth of a standard deviation on the PSI and one-sixth of a standard deviation on the Stanford-Binet.

Covariates are age, race, preschool experience, income, family size, and, for post-test analyses, prescore.

suspicion that where consistent sex-by-model interactions are demonstrated in a variety of sites they may relate not to the "structure" continuum which has come to dominate discussions of preschooling, but to quite different characteristics of models. It is striking, for example, that while the Bushell model seems to favor girls, at least in terms of cognitive outcomes, the other highly-directive model, Engelmann-Becker shows no such tendency. These results are quite consistent with those reported by Bissell (70). In her reanalysis of the Urbana and New York State preschool achievement data she finds effects favoring boys in the one Engelmann-Becker program, but not in another.

The 1969-70 PV data suggest (weakly) that Far West and Tucson programs favor IQ gains of boys over those of girls. If the 70-71 replication supports this hypothesis, I would speculate that this effect might be related to the emphasis both programs (Far West more than Tucson) place on materials and on learning through physical manipulation of objects.[1] Montessori programs, which in Bissell's

---

[1]The 70-71 classroom observation data (the data for 69-70 is not available) lends weight to the guess that Far West and Tucson emphasize learning through physical objects more than most models and that Bushell stresses this less than average. This data gives the frequency of particular types of behavior observed in PV classrooms. The frequency of "adult informing child with concrete object" was consistently higher for Far West sites than for those of any other model (the mean frequency was a shade higher for Weikart, but all the behavior was accounted for by one site. For the other two sites the observed frequency was 0.0); Far West was the only model where some of this behavior was observed in all sites. Bushell sites had the lowest frequency of this observation variable.

The pattern for the variable "child-self-learning with concrete objects" was similar, with only Far West and Tucson showing high frequencies in all sites (the mean frequency per site was higher for Bank Street, but was attributable to only one site; in the other two sites none of this behavior was recorded). Very little of this behavior was recorded in Bushell sites.

analysis (70) show the strongest tendency to favor boys on the Binet, also place a strong emphasis on materials and on children learning through manipulation.

Both the 1969-70 PV data and Bissell's suggest that the interaction of sex and program is likely to be stronger on the Stanford-Binet than on other cognitive tests. In the PV analyses all interaction effects on the PSI are very small (see Table IIIa). In Bissell's analysis regression coefficients for within-model sex effects are far less often significant for the Peabody Picture Vocabulary Test and the Illinois Test of Psycholinguistic Ability than for the Binet. The enormous effects reported by Smith are also on the Binet.

Hypotheses

Strong

1. There will be no consistent main effect of sex on cognitive outcome measures.

   This hypothesis is suggested by previous preschool research, viewed together. It is supported by the 69-70 PV data.

2. Within models the effect of sex on cognitive gains will differ from site to site.

   This hypothesis is suggested by the findings of preschool studies, and supported by the 69-70 PV data.

Weak--Based on 1969-70 data

3. In Tucson and Far West sites IQ differences will favor boys; in Bushell sites IQ differences will favor girls.

   This hypothesis is based on the 69-70 data. The observation data suggests that the tendency of a program to show interaction effects favorable to boys may relate to the emphasis

placed on teaching and learning through the manipulation of concrete objects. Bissell's finding that both Montessori programs in her analysis showed effects favoring boys lends strength to this interpretation.

IV. AGE

Preschool programs are aimed at a population ranging in age from two to six. Most of the experimental efforts based on classroom experiences (as opposed to home-based efforts like Gordon's infant program, or one-to-one tutoring programs like Francis Palmer's) have been directed at children three or older. Although the span of years is short, this is a long time developmentally: a child of three is very different in behavior, demands, and capacities from a child of five and a half. It seems possible that some models of early education would make their greatest impact on children near the bottom of this age range, while others would be most effective with the older children.

Previous Research

Comparative preschool evaluations have not in general been able to explore the possibility of an age-by-program interaction because they have limited the age range of entering children at the start. Having reduced the span of this variable to less than one year, researchers probably assume a priori that it will account for little of the variance.

The assumption that age at entry will be uncorrelated with cognitive outcomes is supported by Hodges, Spicker, and McCandless (67), who, in their comparison of two programs for Appalachian five-year-olds find no significant relation between age and gains. However, their analysis pools groups, examining together children in the experimental

program and those in public kindergarten, so it does not eliminate the possibility of an interaction.

Palmer (70), by contrast, in a concept-training project with two- and three-year-old Black boys, does find a relation between age and cognitive gains, and the suggestion of an age-by-program interaction. Two training procedures were used in one-to-one tutorial sessions; one involved sequenced pre-planned sessions with considerable adult initiation; the other adopted a "discovery" approach, with adults responding to children's initiations, rather than the other way. One year after treatment, results on the principle evaluative measure, the Concepts Familiarity Index, show that for the groups trained at two years of age those taught by the discovery method are significantly above the Training sample, while for children taught at age three the Training group is significantly above the discovery group. On the IQ measure differences between teaching methods are insignificant, but age of training does appear to make a difference: threes outperform twos immediately after training and one year later.

Van de Riet's (72) evaluation of Sprigle's "Learning-to-Learn" program also shows important effects of age on cognitive gains: the mean first-year IQ gains for children entering the experimental program at age four is twice that of children entering at age five (9.1 points vs. 19.7 points), even though the initial IQ of the two groups is the same.

These findings do not provide specific hypotheses concerning age-by-model interactions, but they do suggest that, within particular programs, age may be strongly related to gains.

1969-70 Data

When the PV sample is divided at the median age of entry (59 months in October) into older and younger groups, a few models include only a few children in one group or the other. This is because in particular sites the age range is narrow--usually about one year. Age, thus defined, is therefore confounded with site and sponsor, so an analysis of variance model is an inefficient way of looking for main effects and interactions based on age.

Since age is, in any case, a continuous variable, regression analysis has been used instead of ANOVA. For each model age is regressed against post-test score (gains are not used on these analyses) on both the PSI and the Stanford-Binet, controlling for prescore, sex, previous preschool experience, race, race-by-sex interaction, mother's education, income, and family size. The magnitude and direction of the regression coefficients so obtained gives a measure of the effect of age on the cognitive post-test score. Since the unit of age is months, a coefficient of .5 means that a child of five has an advantage amounting to 6.0 points on the post-test over a similar child who is a year younger. Similarly, a coefficient of -1.0 would mean that for the model in question four-year-olds outperformed comparable five-year-olds by an average of 12.0 points.

The partial confounding of age span and model raises some question about the validity of this approach. Specifically, if the effect of age on post-test is not linear, then the regression coefficients for programs with mainly older or mainly younger children will show bias. To check this possibility, age coefficients were calculated separately

for older and younger children for those models with a sufficient number of each.[1] These analyses revealed no significant differences between age coefficients for older and younger children within models. The effect of age on post-test is therefore assumed to be linear within models.

Preliminary analysis of PSI data for the entire sample (1408 children) revealed an interaction between age and previous preschool experience: the age coefficients for children with and without preschooling were significantly different from one another. For children with preschool experience, B = .0097 (n = 472); for those without such experience, B = .205 (n = 936). The first value is insignificant while the second is significant above the .005 level. For this reason separate age coefficients were obtained[2] for children with and without preschool experience both for the PSI and for the Stanford-Binet. For the PSI, coefficients were calculated for the whole sample and then separately, where possible, for children under 60 months. This was done because the distribution of post-test scores indicated the possibility of a slight ceiling effect for older children with prior preschool experience. Although this effect was not strong, it might have been expected to bias downward the coefficient for programs which mainly served this group.

---

[1]This analysis was carried out for the IQ only. The procedure for the PSI is described below.

[2]Separate regression equations were calculated for each model, using as independent variables age, pre-test score, sex, race, race-by-sex interaction, preschooling, mother's education, family size, and income.

Part I

Table IVa

PSI Post Test
Regression Coefficients for Age (in months) for Children
without Previous Preschool Experience*
1969-70 PV Data

		Tucson	Bushell	Gordon	Bank St.	Far West	Weikart	EDC	E-B
All PV:	N	77	8	53	166	27	116	52	25
	B	.901	.315	.319	.202	-.0766	.066	-.028	-.201
Young only:**									
	B	.676	.603	.550	.353	.024	-.050	-2.10	

*Other variables in equations include pretest score, sex, race, race-by-sex interaction, previous preschool experience, mother's education, family size, and income.

**Under 59 months of age in October

For children without preschool experience, the pattern of coefficients is not altered by including older children in the analysis. Either way, the regression coefficients obtained for Tucson, Bank Street, Bushell and Gordon programs exceed those obtained for the whole sample (.205), while those obtained for Far West, Weikart and EDC are less than .205.[1] Between-model differences are not significant[2] for young children without prior preschooling; when older children are included, Tucson is significantly different from (greater than)[3] all groups except Bushell.

For children <u>with</u> previous preschool experience the pattern is reversed. For the younger sample the coefficient for age is <u>less</u> than .0097 for Tucson, Bank Street, and Bushell programs and greater than .0097 for Weikart, Gordon and EDC.[4] When older children are included more programs show negative coefficients, suggesting a ceiling on the post-test for older children with previous preschooling.

---

[1]Becker-Engelmann has too few young children to be included in the "young only" analysis; in the full-sample analysis it falls with Far West, EDC, and Weikart.

[2]Two-tailed T-tests were used to evaluate the significance of differences between B-weights. Since separate equations were used for each model this test is not strictly valid statistically. It is a heuristic indicating the magnitude of the effect.

$$T = \frac{B_1 - B_2}{\sqrt{SE_1^2 + SE_2^2}}$$

[3]Becker-Engelmann and Far West do not have enough young children with preschool experience to be included.

[4]The coefficient for age is observed to be greater in Tucson than in other models. A two-tailed test shows it to be significantly greater.

Within the younger sample the age coefficients for Weikart and EDC programs are significantly different from (greater than) those for Bank Street, Tucson and Bushell. The Gordon program is not significantly different from any of the others. When the whole sample is considered the only significant difference is between Tucson and other programs.

These coefficients represent in some cases a fairly substantial effect. Thus, in the Tucson program an age increase of one year for children without earlier preschooling is worth more than ten points on the PSI post-test, with other variables controlled. For the Becker-Engelmann program the difference is in the opposite direction, with younger children scoring somewhat higher than older ones on the post-test, after the effect of the other variables is taken into account.

The data on IQ for children without preschool experience shows a pattern very similar to the PSI: for Far West, Bank Street, Engelmann-Becker, Weikart and EDC the relationship between age and adjusted post-test is negative. For Tucson, Bushell[1] and Gordon it is positive. T-tests for differences among the coefficients show the coefficient for Tucson to be significantly less than that for Bushell and greater than those for other programs. The differences between coefficients for Gordon, Weikart, Far West, EDC, Becker-Engelmann and Bank Street are insignificant.

---

[1]Results for the Bushell program should be cautiously interpreted since the n is very small.

Part I

Table IVb

PSI Post-Test
Regression Coefficients for Age (in months) for
"Young"* Children with Preschool Experience**
1969-70 PV Data

Tucson	Bushell	Gordon	Bank St.	Weikart	EDC
-1.43	-.349	.510	-2.36	1.28	2.06

*Under 59 months in October.

**Other variables in equations include pre-test score, sex, race, race-by-sex interaction, previous preschool experience, mother's education, family size, and income.

Part I

Table IVc

IQ Post-Test
Regression Coefficients for Age in Months*
1969-70 PV Data

	Tucson	Bushell	Gordon	Bank St.	Far West	Weikart	EDC	E-B
No Previous Preschooling	1.0731	4.206	.148	-.2520	-.8864	-.177	-.144	-1.550
Previous Preschooling	---	---	-.037	---	---	-.389	.209	-.4919
N	47	14	34	48	17	43	45	61

*Other variables in the equation include pre-test score, sex, race-by-sex interaction, previous preschool experience, mother's education, family size, and income.

IQ data for children with previous preschooling are available for only a few programs (Engelmann-Becker, Weikart, Gordon and EDC). Coefficients for all these programs are insignificant; they are negative for three of the four.

These data on the relationship between age and post-test on two cognitive tests are very far from providing a clear picture. However, some rather murky patterns do emerge. First, the relationship between age and gains appears to be stronger for children without previous preschooling than for those with it. Coefficients are larger, and for the sample as a whole (sponsored and unsponsored children) the age coefficient for children without prior preschooling is significant while that for children with previous preschool is not.

Second, for children without previous preschooling, certain programs seem to promote cognitive growth more effectively for older children while other programs achieve their best results with younger children. These age effects are consistent for PSI and IQ: younger children gain more than older[1] in Far West, Becker-Engelmann, EDC and Weikart programs; older children gain more in Bushell, Tucson and Gordon programs. For Bank Street, results for the two tests are inconsistent, with age showing a significant positve effect on PSI score and an insignificant negative effect on IQ.

Third, the relationship between age and cognitive gains within a program does not appear to be related to the degree to which the

---

[1] In the case of the PSI this means that younger children gain more than expected with relation to older children, since the cut-off is .205, not zero. This is because the PSI, which is not standardized by age, shows positive effect for age.

program could be considered "directive." Neither does it relate to other obvious attributes of the models.

Hypotheses

Weak--Based on 1969-70 PV Data

1. The relationship between age and adjusted post-score on PSI and IQ tests will be stronger and more significant for children without previous preschool experience than for children with it.
2. For children without previous preschooling, age will be positively related to IQ and PSI post-test in Tucson and Bushell sites. In other models the relationship between age and post-test score will be negative or very weakly positive.

V. SOCIO-ECONOMIC STATUS

Much of the work on experimental preschool approaches which was done during the sixties was based either implicitly or explicitly on the idea that a program which was satisfactory for middle-income children was not necessarily optimum for low-income children. The traditional nursery school has been the preschool experience of the middle class for over a generation; it was hoped that some other approaches might yield more impressive results with the disadvantaged. In large part, of course, the new approaches grew out of new objectives for preschooling--cognitive preparation for school, rather than socio-emotional development; however, the idea that different strategies are appropriate to different populations was often also implicit.

It seems therefore important to ask, as a good many researchers have done before, whether, within the PV sample of Head Start models, some approaches are most successful with the most disadvantaged and others with the least disadvantaged. Clearly, the PV sample of children is not optimum for pursuing this question, since all children who are eligible for Head Start come from relatively poor families. However, it is still of interest to know whether within the Head Start target population SES is differently related to gains in different programs.

Previous Research

A number of researchers have explored the possibility of an interaction between SES and programs for preschool children. Most recently

Bissell (70) has reviewed previous studies and re-analyzed the data collected by Karnes, DiLorenzo and Weikart in Urbana, New York State and Ypsilanti. Her results, and those of other researchers, suggest that cognitive gains are differently related to SES in different programs. Specifically, Bissell finds that in less-structured models (traditional, Montessori, and Weikart), gains are positively related to SES, while for highly-structured programs (Bereiter-Engelmann and Karnes Amelioration Program), the relationship is either more weakly positive or negative. These findings are compatible with the earlier findings of Sprigle and Van de Riet and of most others who have explored the SES-by-model interaction.

The data of Kraft, Fuschillo and Herzog (68) extend this finding in an interesting way, raising the possibility of a second-order interaction between SES, program, and previous preschooling. As noted in the section on previous preschool experience, this study of a traditional nursery school finds that while SES is related positively to gains for the first year of this program, the relationship is negative for the second year.

1969-70 PV Data

Using a measure which weights equally income, mother's education and family size, children are divided into two SES categories, "high" and "low".[1] Four analyses of covariance (IQ and PSI gains and

---

[1]Income, mother's education, and family size are all standardized; each variable is given a mean of 0.0 and a S.D. of ±1.0. A child's "SES score" is the sum of his standardized scores on these three variables. If his SES score is above the mean he is assigned to the "high" category; if it is below the mean he is assigned to the "low" group. For the 315 children included in the IQ analyses sample means are as follows: income, $3401; mother's education, eighth grade; family size, 5.5.

post-test) have been performed for PV children in sponsored classrooms who had taken both IQ and PSI tests. Children are stratified by SES category and by model; the main effect of SES and the SES-by-model interaction effect have been computed.

Main effects of SES are significant, though small, for both IQ analyses, favoring high SES children (the difference between adjusted means being 2.74 for IQ post-test and 2.662 for IQ gain). Main effects for the PSI are in the opposite direction and insignificant.

Interaction effects are small and insignificant overall for all analyses. The patterns of the adjusted effects on these analyses, such as they are, are remarkably consistent with one another (see Table Va). In general, the four analyses show weak effects favoring low SES children in Bank Street, Far West, Tucson and EDC, and favoring high SES children in Weikart, Bushell, and Engelmann-Becker. These results are inconsistent with Bissell's[1] in all respects save one: in her analysis and in the PV data, Weikart's program shows an effect favoring high SES children.

It at first seemed possible that these puzzling results might be accounted for by the two-way interaction involving previous preschooling which the data of Kraft, Fuschillo and Herzog suggested. However, a second set of analyses using only children without previous preschooling reveals a similar pattern: overall interaction effects

---

[1]The PV sample of models is somewhat different from Bissell's: specifically, models in the middle range of "directiveness" are somewhat more heavily represented.

Part I

Table Va*

Interaction of SES Category with Model
Analysis of Covariance:
Adjusted Interaction Effects Greater Than 1.0**
1969-70 PV Data

	Positive for High SES	Positive for Low SES
PSI Post	Weikart (±1.088)	Far West (±1.999)
PSI Gain		Bank St. (±1.061)
IQ Post	Weikart (±2.999)	Far West (±1.413) Bank St. (±1.821)
IQ Gain	Weikart (±1.575)	Far West (±1.026) Bank St. (±1.179)

*All analyses include all models.

**This means that with the main effect of SES category taken out the difference between adjusted means is greater than 2.0 points. This amounts to one-fifth of a standard deviation on the PSI and one-sixth of a standard deviation on the Stanford-Binet.

Covariates are age, sex, race, preschool experience, and, for post-test analyses, pre-test score. See page 32 for explanation of adjusted effects.

are insignificant, but the Weikart programs show a fairly strong effect favoring high SES children, while the Bank Street program showed consistent effects favoring low SES children.

These results raise rather than settle questions. Without the PV data, there would have been every reason for a hypothesis that more directive programs favor low SES children and less directive models favor high SES children. The 1969-70 analyses favor an opposite hypothesis. I cannot suggest any totally satisfactory resolution of this contradiction, but two points seem pertinent. First, because the PV sample is a national one, our indices of SES may be inadequate: a given income, educational level, and family size may add up to quite different SES levels in Mississippi and Duluth; by pooling children in different regions we are probably diluting the strength of our SES measure. Unfortunately we would have to know more than we do about economic and social conditions in these Head Start communities in order to make appropriate regional adjustments.

Second, we know that programs do not look exactly the same in Planned Variation classrooms as they do in sponsors' experimental preschools (Lukas and Wohlleb, 72). It may be that the interactions we observe (such as they are) result not from intrinsic properties of the PV models, but from a complex interaction involving the demands of the model, its ease of implementation, and the expectations of teachers, administrators, and parents in Head Start. Thus, teachers attempting to implement models which emphasize cognitive development may inadvertently concentrate on the most receptive children--those of high SES. Teachers implementing models which are closer to their

own experience and make fewer demands for cognitive gains may feel freer to concentrate on the children who seem to need help the most. Head Start itself, as an institution, may frame teacher responses in distinctive ways.

The sensible reaction to these puzzles is to take a hard look at the 1970-71 data. In the meantime, the first year's data raise the suspicion that interactions between SES and model within the institutional and local framework of Head Start may be quite different from what one gets in experimental preschools. If the 1970-71 data confirm the pattern of the 1969-70 data, then we will have to explore the reasons for this at greater length.

Hypotheses

Strong

1. Within-model effects will favor high SES children in the Weikart model.

   This hypothesis is suggested by Bissell's analysis and supported by the 1969-70 PV data.

Weak--Based on 1969-70 PV Data

2. The overall interaction between SES and model will be insignificant both for the PSI and for the Stanford-Binet.
3. Within-model interaction effects on the Stanford-Binet will favor low SES children in the Bank Street and Far West programs.

VI. ETHNICITY

Only two ethnic-cultural groups are represented in sufficient numbers in the PV sample to be included in an ethnicity-by-model analysis. Indian and Spanish-speaking groups are included in too few sites, and in insufficient numbers, to make a cross-program comparison possible. For this reason, the present analysis is limited to Black and white children, and to those few models which include enough of both to make within-model comparisons possible.

Previous Research

Two of the studies reviewed by Bissell (70)--DiLorenzo (69) and Teska (69)--examine the relative performance of Black and white children in a range of programs. Like Bissell, these researchers report greater Stanford-Binet gains for whites than for Blacks across programs. However, neither DiLorenzo nor Teska control for SES, so these findings are predictable (gains are usually correlated with SES; Blacks are generally of lower SES than whites). Bissell does control for SES. Using the same SES scale for both races she finds whites gaining more than Blacks of comparable SES. When the SES scores of Blacks are adjusted downward (on the assumption that the actual status of a Black family and a white family with identical SES ratings are different) Bissell finds no clear pattern in the achievement of Black and white children. Erickson (69), in his comparison of traditional and Bereiter-Engelmann approaches, finds no evidence for a race-by-program interaction.

1969-70 PV Data

Only five of the eight models--Tucson, Engelmann-Becker, Bushell, Gordon and EDC--are included in the PSI analyses: the others have very few children from one of the two ethnic groups. On the IQ analyses the number of models is further reduced by elimination of Gordon and Engelmann-Becker (models are excluded from an analysis when one includes fewer than seven children).

The main effect of race is significant in only one of the four analyses of covariance--the PSI post-test. This analysis shows a small effect favoring white children--a difference of 2.2 points on the covaried post-test score. In the two gains analyses, effects are insignificant, but favor Black children. On both tests, Black children pre-test lower and gain somewhat more than white children.

Interaction effects are significant on all four analyses (see Table VIa). The strongest consistent interaction effects are those favoring Black children in the Bushell model and white children in Tucson and Engelmann-Becker.[1] I know of no logical way in which to interpret these results; they lend weight to the suggestion that consistent, interpretable race-by-model interactions are unlikely to emerge. Table VIa suggests that the dimension of directiveness is irrelevant to any interactions observed. If the effects suggested in these analyses are replicated in the 1970-71 data, it would be in order to speculate on the reasons for the marked difference between Bushell and Engelmann-Becker.

---

[1]The pattern observed for EDC reflects site effects: the mainly Black EDC site in Washington, D.C. shows larger PSI gains than the mainly white site in Johnston County; the Johnston County site, on the other hand, shows greater IQ gains.

Part I

Table VIa*

Interaction of Ethnicity with Model
Analysis of Covariance:
Interaction Effects Greater Than 1.0 Points**
1969-70 PV Data

	Effects Favoring Black Children	Effects Favoring White Children
PSI Post	Bushell (±1.081) Gordon (±3.309)	Tucson (±3.168) Engelmann-Becker (±2.109)
PSI Gain	Bushell (±2.570) EDC (±1.103)	Tucson (±1.920) Engelmann-Becker (±2.750)
IQ Post	Bushell (±3.005)	EDC (±2.864)
IQ Gain	Bushell (±3.010)	EDC (±3.695)

*PSI analyses include Tucson, Engelmann-Becker, Bushell, Gordon, and EDC. They exclude Far West, Bank Street, and Weikart. IQ analyses include only Tucson, Bushell, and EDC.

**Covariates are age, sex, preschool experience, income, family size, and, for post-test analyses, pre-score. For an explanation of adjusted effects, see page 32.

Hypotheses

Weak--Based on 1969-70 Data

1. Within-program interaction effects will favor white children in Engelmann-Becker and Tucson programs and Black children in Bushell programs.

VII. RESPONSE STYLE

Response style refers to the way in which an individual behaves when confronted with an intellectual demand--to the manner in which he characteristically attacks or avoids cognitive problems. Conceptually, it is independent of the correctness of a child's answers to particular questions, although it is of course possible--depending on how response style is defined--that bright children will adopt particular styles more often than dull children.[1] It seems reasonable to suppose that certain educational approaches and assumptions would prove effective with children who characteristically deal in a particular way with cognitive demands, while others might work better for children who employ a different style.

The Planned Variations data provides one interesting clue to a child's response style: the Hertzig-Birch scoring of the Stanford-Binet. Instead of simply marking children's responses to each item right or wrong, testers used one of the eight Hertzig-Birch categories to classify his answer. The following selection from the Binet testers' manual describes the codes. The first two categories describe correct answers. The last six are used for incorrect ones.

---

[1] In this study "response style" is defined by the Hertzig-Birch scoring of the Stanford-Binet. The response style variables derived from this instrument do not, in fact, correlate strongly with initial IQ. See Appendix A. For analysis of response style data on two different populations see Hertzig, M.E., Birch, G., Momas, A., and Mendez, O.A., "Class and Ethnic Differences in the Responsiveness of Preschool Children to Cognitive Demands", Monographs of the Society for Research in Child Development, 33 (1), 1968.

Coding Categories

1. Delimitation. The child's response to a work item answers the item but provides no further elaboration. For example, a child might correctly fold his paper to match that of the examiner and then sit quietly or give the response, "wood", to the question, "What is a house made of?"

2. Spontaneous Extension. The child work response is accompanied by an unsolicited elaboration related to the item. For example, after she finished stringing beads, a girl ties the ends of the string together and tries to slip it over her head, or she might say, "Yours is smaller than mine", in comparing her tower of four blocks with the examiner's model consisting of three blocks.

3. Incomplete. The child fails to complete the task (either a verbal or non-verbal task) and does nothing else (categories 4, 5, 6, 7 below).

4. Negation. Direct refusal to work, such as "No, I won't", "I don't want to", or "I don't like to do it" or shakes her head or turns away to indicate refusal.

5. Substitution. The child offers an irrelevant verbalization or engages in irrelevant physical activity instead of responding to the task requested. For example, a substitute verbalization often takes the form of a request for an alternative activity, such as: "I want to play with the toys now", "I want to go to mommy", or "I want a drink". Non-verbal substitutions may be of the following type: When asked to build a block bridge, the child gets up, goes to the toy shelves and begins to play with a truck. When asked to describe the pictures, the child gets up and runs out of the room.

6. Competence. The child states some limitation of ability to perform the assigned task. Such responses include the following: "I don't know how", "I'm too little to do it". It is possible, though unlikely, that the child can convey his feeling of lack of competence by use of gestures and animation.

7. Aid. The child makes a direct request for help from the examiner. This would include such comments as, "Show me how to do it", or "Tell me what the answer is". It is unlikely that a request for aid will be made non-verbally.

8. Passive. This is a No Response category. The child may just sit still when, for example, sticks are presented, or look straight ahead and say nothing when asked to tell a story about pictures.

These categories were originally worked out in a study comparing the cognitive styles of lower class Puerto Rican and middle class white children in New York City (Hertzig et al., 1968). For this study the entire stream of each child's behavior was described; the categories were developed empirically from the data.

The procedure used by the Planned Variations study differs in two ways from that of the Hertzig study. First, the PV tester codes only the last response to each item. This is necessary for reasons of reliability, since testing and coding are done by the same person (in the Hertzig study one person administered the test while another recorded the child's behavior). We do not know how much information is lost in this way.[1] Second, the categories used by PV describe incorrect work responses less completely than those of the original study. Categories 1 and 2, delimitation and extension, apply only to correct answers; there is no equivalent distinction for an incorrect response. Since testers categorize most incorrect responses as "incomplete" (code 3), information is certainly lost.

Despite these limitations the Hertzig-Birch scoring may provide interesting information about a child's way of dealing with a set of cognitive problems which range from the very easy to the very difficult. It seems intuitively likely that a child who often responds to difficult items by substituting a different activity may

---

[1]The loss could be considerable. Thus, for example, if a child works for eight minutes on the block bridge and then looks up at the tester, saying, "I can't do it", his response is described only by code 6. The researcher has no way of knowing whether the child attempted the problem at all.

behave differently in the classroom than a child who becomes mute and unresponsive when faced with problems he cannot solve. Even though the Hertzig-Birch coding does not provide a total picture of the child's behavior, it may give a valuable clue to his style, a clue which would help us to predict the type of program in which he will learn best.

Previous Research

Unfortunately, there is almost no research which aids in predicting the nature or strength of an interaction between "style"--as indicated by the Hertzig-Birch data--and model. Predicting, or even understanding, such an interaction is peculiarly complex, I think, because we know so little about what the behaviors indicate--or how they might interact with characteristics of models.

The behavior so coded may provide two kinds of information about a child. The first concerns his state of mind at the time of testing: a particular response, say passivity, may be an indication of anxiety, boredom, fear of failure, or confusion. If we had solid evidence that such a response was characteristically associated with one of these (for example, test anxiety) we might have some basis for predicting an interaction between frequency of passive responses and gains in particular programs.

There is, however, a second aspect to the situation. A particular behavior is important not only as a sign of the child's feeling about testing. It is important in itself. Teachers probably respond differently to children who meet their demands with irrelevant

verbalizations, whatever the reason for this behavior. Their response--and the child's subsequent learning and behavior--may vary from model to model. (It will certainly vary from teacher to teacher.)

There is little research to aid us either in deciding what particular types of responses might "mean"--what state of mind on the part of the child they usually indicate--or how aspects of different models might lead adults to respond differently to the behaviors themselves. A study of the behavior of primary school children during Rorschach testing (Sarason, 58) suggests that two of the Birch behaviors, Negation (code 4) and Substitution (code 5) might be related to test-anxiety.[1] However, the 1969-70 data indicates that the variables derived from these two codes do not interact strongly with model characteristics. For "passive" and "competence" responses, which do show some interaction with model, we have little beyond intuition to illuminate the meaning of the behavior.

---

[1]In this study, sixty-four children who had previously been rated on test-anxiety according to their answers to a questionnaire were given Rorschach tests. Descriptions of their test-taking behavior were submitted to clinicians, who were asked to classify each child as high or low on test-anxiety (of the 64 children, 32 were in the top quartile of test anxiety while the other 34 were in the bottom quartile). The clinicians were asked to indicate which behaviors they had used to classify children, and the frequencies of these behaviors among high and low anxiety children were then tabulated. Two which might relate to our categories and were positively related to test-anxiety were "rejection of one or more cards" ($p = .05$) and "not responding to the stimulus area of the card" ($p = .025$). However, one would clearly have to be tentative in making a parallel between Rorschach and Binet testing.

1969-70 Data

Lacking any persuasive theoretical grounds for predicting particular interactions between response style and PV models, it seems essential to take an empirical approach to the data. The frequency distributions (Table VIIa) reveal that both in the fall and in the spring over 85% of all responses are coded 1 (delimitation) or 3 (incomplete). When answering correctly, children do not usually go beyond the requirements of the task; if unable to answer correctly they still generally make a relevant "work" response. Table VIIa gives mean and median frequencies of each response (per child) for fall and spring testing, plus the percentage of children making no response in that category.

The response style analyses are aimed at discovering how specific deviations from the usual response pattern (a high incidence, for example, of extensions or passive responses) relate to cognitive gains in particular models. The Hertzig-Birch codes are of interest for what they reveal about initial differences between children--not differences which result from participation in particular programs. For this reason only responses on the IQ pre-test are used. Since the test was given about three weeks after the opening of school it is not inconceivable that children's manner of response might already have been affected by different model emphases. If this were true it would complicate interpretation of any interactions

Part I

Table VIIa*

Frequencies Per Child of Each Hertzig-Birch Code for Children with Valid Fall and Spring IQ Tests (n = 315)
1969-70 PV Data

		Fall			Spring		
Code	Label**	% of children making 0.0 responses	Mean	Median	% of children making 0.0 responses	Mean	Median
1	Delimitation	0	66.381	64.000	0	62.41	59.0
2	Extension	51.75	2.473	0.0	68.57	.940	0.0
3-8	Incorrect	.63	49.263	47.000	0.0	50.997	50.000
3	Incomplete	0	37.965	36.000	5.08	38.651	38.000
4	Negation	89.84	.270	0.0	90.48	.270	0.0
5	Substitution	43.17	3.165	1.0	75.56	1.044	3.332
6	Competence	30.79	4.717	3.0	33.33	7.797	3.000
7	Aid	90.79	.162	0.0	87.94	.276	.968
8	Passive	52.38	2.984	0.0	60.95	2.959	0.0

*Sample: all children with valid scores on both Stanford-Binet and PSI, fall and spring.

**See page 81 for description of categories.

observed. In fact, however, it does not appear to pose a serious problem.[1] There seems little object in examining categories which showed minimal variability, so codes 7 (request for aid) and 4 (refusal) are eliminated. Only 10% of the children have any responses at all which are coded in those two categories at pretest. The remaining four categories--extension, substitution, competence, and passivity--are not highly correlated either with one another, or with background variables (See Appendix A for intercorrelations among the four variables and their correlations with age, preschool experience, IQ, sex, race and SES). It therefore seems logical to examine each one separately. Children in all models who have valid scores on the Stanford-Binet pre- and post-tests are characterized as "high" or "low" on each of the four behaviors depending on how many responses in each category they made on the initial IQ test. For each variable, the pretest sample median is taken as the cut-off between high and low (for extensions and passive responses low = 0.0, for substitutions

---

[1] To check this possibility I have looked at the number of children in each model scoring above and below the sample median for "competence" and "passive" responses (these are the Hertzig-Birch variables which show some sign of interacting with model--see text). A x^2 test shows no significant differences between models for "competence" responses--indicating that program emphases do not significantly affect children's tendency to use this type of response by the time of pretest. Between-model differences are significant ($p < .05$) for passive responses. However, the significance of these differences is attributable to the high proportion of children in the Gordon model making no passive responses at all. Since this variable (frequency of passive responses) does not appear to interact strongly with the Gordon model, between-model differences do not, in my mind, pose a major problem. They are, however, explored further in Appendix B.

low < 2.0; for competence low < 4.0).[1] Two-way analyses of covariance have been performed for each variable, using as dependent variables PSI and IQ post-test and gains. The analyses are done by model and frequency of behavior, using the categories high and low.[2]

A. Passive Responses

A response is coded "passive" when the child made no response, verbal or otherwise, to the item. Children are considered "high" on this variable if they make any passive responses at all (the median for the sample being 0.0).

On all four analyses of covariance--PSI gain, PSI post-test, IQ gain, IQ post-test--the main effect of passive response is insignificant. In effect this means that when all programs were considered together the covaried post-test scores and gains of children making no passive responses do not differ significantly from those of children making one or more such responses.

The interaction effects vary in magnitude according to the analysis; the significance of the overall interaction ranges from

---

[1]The question arose as to whether using raw frequencies of a code, rather than percentage of correct or incorrect answers, would distort the analysis. The argument was that brighter children would have more right answers and fewer wrong ones than dull children and would, in consequence, have more opportunities to "extend" and fewer occasions for substitutions, passive responses, etc. This is not in fact the case, because each child, regardless of his IQ, starts at the level where he passes all tests and stops at the level where he fails all tests. In consequence, the correlation between number of items correct and IQ is .051, which is insignificant (correlations done for entire PV IQ sample; n = 315).

[2]The sample for these analyses consists of children for whom the Planned Variation study has valid pre- and post-test data for both the PSI and the Stanford-Binet; n = 305).

.163 for the covaried IQ post score to .067 for IQ gain.

We might predict from Table VIIb that Far West, Tucson, and Weikart will show greater cognitive gains with children low in passive responses, while Engelmann-Becker, Bushell, and Bank Street would show positive effects for those high in passives. Once again, the observed interaction pattern does not follow the directiveness continuum strictly, but the two behaviorist programs do fall together, both favoring the gains of children high in passive responses.

B. Competence

The code "competence" is used for any response by which the child indicates he is unable to solve the problem. Ordinarily this means saying "I don't know", or "I can't", although there are other ways of communicating the same message. The frequency of such responses increases sharply from fall to spring testing (the mean number per child going from 4.7 to 7.8). Interestingly, this change is not in the number of children making any such responses (69% in the fall and 66% in the spring) but in the number using this response quite frequently; the percentage of children making more than 10 such responses rises from 11.7% to 24.5%. One might guess from this data that "competence" responses are not a strategy one learns in preschool, but one which, if already mastered, gets frequently exercised there. The main effect of "competence" responses is significant for both IQ analyses, favoring children who make many such responses. It is insignificant for the PSI analyses.

The evidence for an interaction between frequency of "competence"

Part I

Table VIIb*

Interaction of Frequency of Passive Responses with Model Analysis of Covariance: Adjusted Effects Greater Than 1.0**
1969-70 PV Data

Analysis	Positive effect for children low in passive responses (no passive responses)		Positive effect for children high in passive responses (one or more passive responses)	
PSI Gain	Far West Tucson Weikart EDC	±1.657 ±1.471 ±2.745 ±1.885	Bank St. Engelmann-Becker Bushell	±2.582 ±1.869 ±2.567
PSI Post	Weikart EDC	±1.795 ±2.028	Bank St. Engelmann-Becker	±2.744 ±1.231
IQ Gain	Far West Tucson	±4.844 ±1.025	Bushell EDC	±3.750 ±1.317
IQ Post	Far West	±4.017	Gordon	±2.418

*All sites included in all four analyses.

**Covariates included sex, race, preschool experience, family size, income, and age. Pre-score was included for the two post-test analyses, but not for gains analyses. See page 32 for explanation of adjusted effects.

response and cognitive gains in particular models is neither strong nor consistent. The overall interaction effect is insignificant for all analyses (the closest approach to significance is in the IQ gain and PSI post analyses; $p = .093$ for one and .125 for the other). Furthermore, the within-program interaction effects are not totally consistent for the four analyses (see Table VIIc).

The pattern of effects, though somewhat mixed, suggests that the "directiveness" continuum might be relevant to interactions of "competence" with model: Engelmann-Becker shows effects favoring those low in competence responses while Bank Street appears to benefit those high in such responses. Although interactions of EDC and Far West with this variable do not meet our criteria for proposing a specific hypothesis for the two models, both show signs of favoring children high in competence responses. For Bushell, Tucson and Weikart models different analyses show contradictory patterns.

C. Substitutions and Extensions

A response is coded "substitution" when a child substitutes an activity of his own choosing for the problem posed by the tester. It is coded "extension" if, after responding correctly, he goes beyond the requirements of the task either verbally or non-verbally. All analyses of covariance done for these two variables show highly insignificant main effects and interaction effects. The magnitude of such interactions as do appear is small enough to make further consideration of either variable appear pointless.

Part I

Table VIIc*

Interaction of Frequency of Competence Responses with Model
Analysis of Covariance: Adjusted Effects Greater Than 1.0**
1969-70 PV Data

Analysis	Positive for those low in competence responses (0-3 competence responses)		Positive for those high in competence responses (more than 4 competence responses)	
PSI Gain	Engelmann-Becker Bushell	±1.497 ±2.797	Weikart Far West Tucson Bank St.	±1.083 ±1.794 ±1.313 ±1.470
PSI Post	Far West Engelmann-Becker	±2.402 ±1.546	Bank St. EDC	±2.449 ±1.234
IQ Gain	Tucson Weikart	±2.049 ±1.101	Far West Gordon	±5.114 ±1.350
IQ Post	Tucson Weikart	±1.676 ±1.551	Far West Bushell	±1.255 ±1.774

*All sites included in all four analyses.

**Covariates included age, sex, race, preschool experience, income, and family size. See page 32 for explanation of adjusted effects.

D. Discussion

The results of the analyses of covariance for the four child variables derived from Hertzig-Birch codes suggest that two of the variables--frequency of passive response and frequency of competence response--may interact with model characteristics, and that the two others almost certainly do not.

The data suggests--weakly--that, both for passive and for competence responses, the model-by-child variable interaction may relate to the "directiveness" continuum. Specifically, children high in competence responses and/or low in passive responses appear to do better in less-directive models (EDC, Far West, Tucson) while, conversely, those low in competence responses and/or high in passives do somewhat better in the more-structured models (Engelmann-Becker and Bushell). This formulation is not hard and fast (Bank Street falls with Engelmann-Becker and Bushell in favoring those low in passive responses), but it appears to be of some use.

These patterns make intuitive sense: it would seem that both these responses indicate something about a child's capacity--and willingness--to deal with cognitive problems that he finds difficult. Saying "I don't know", or in some other way directly acknowledging ignorance, seems to be a fairly competent response; it is also a response which will often result in adults supplying the need information and skills. Complete passivity seems, on the other hand, to be a less competent response and, in a classroom where adults as well as children have a good many pedagogical choices, one that may less frequently lead to learning. For these reasons we might expect

children low in passives and high in competence responses to utilize adult resources relatively well in less-directive classrooms. This skill might be less of an asset in a more-directive classroom and, indeed, the children least able to mobilize adult resources on their own might benefit most from a highly-directive approach.

Another consideration which may be important is the degree to which a model provides guidelines for dealing with the particular behavior. Bank Street may show an interaction favoring children high in passive responses because of a particular model emphasis on withdrawn, unresponsive children. Where the model acts to focus teacher's attention on a particular group of children it may benefit these children unexpectedly.

Hypotheses

Weak--Based on 1969-70 Data

1. Within-model interactions will favor children making few competence responses in Engelmann-Becker programs. In Bank Street programs, interactions will favor those making many such competence responses. More generally, in more directive programs interaction effects will favor children making few competence responses, while in less-directive models they will favor children making many such responses.
2. Within-model interactions will favor children making many passive responses in Bank Street, Bushell and Engelmann-Becker programs; in Weikart, Tucson and Far West programs interaction effects will favor those making few passive responses.

VIII. SUMMARY OF PART I AND GENERAL PREDICTIONS

Specific Interaction Hypotheses

The 1969-70 Planned Variation data, in conjunction with selected preschool studies has suggested the following hypotheses about specific interactions of child characteristics and preschool model which we might expect to observe. These hypotheses have been discussed in the preceding sections. A summary list is given below.

Initial Ability

Strong

1. Less-directive models will show interactions favoring children of high intial IQ on the PSI post-test.
2. Weikart programs will show a stronger tendency than other programs to increase the variance of IQ scores.

Weak - Based on 1969-70 Data

3. In Gordon programs within-model interaction effects on the PSI will favor children of low initial IQ. The variance of IQ scores will tend to decrease in this model.

Prior Preschooling

Strong

1. The cognitive gains of children having previous preschool experience will be smaller than those of children having no such experience.
2. Children with a particular type of preschool experience will make greater second-year IQ gains in a program which

is different from their first experience than in one which is similar to it.

3. Children whose first preschool experience was in a highly-structured program will be more likely to maintain or increase IQ gains during the second preschool year than children whose first-year program was less structured.
4. Children making large IQ gains in their first preschool year will make smaller gains during the second year than children making negligible gains the first year.
5. Children of low initial IQ will make greater gains during the second year of preschooling than children of high initial IQ.
6. Children of low SES will make greater second year IQ gains than those of high SES.

Weak--Based on 1969-70 Data

7. In Tucson and Bank Street programs, interaction effects on both cognitive measures will favor children with previous preschooling; in Gordon programs, interaction effects will favor children without such experience. In Engelmann-Becker programs, interaction effects on the PSI will favor those without previous preschool experience.

Sex

Strong

1. There will be no consistent main effect of sex on cognitive outcome measures.

2. Within models the effect of sex on cognitive gains will differ from site to site.

Weak--Based on 1969-70 Data

3. In Tucson and Far West sites IQ differences will favor boys; in Bushell sites IQ differences will favor girls.

Age

Weak--Based on 1969-70 Data

1. The relationship between age and adjusted post-score on PSI and IQ tests will be stronger and more significant for children without previous preschool experience than for children with it.
2. For children without previous preschooling, age will be positively related to IQ and PSI post-test in Tucson and Bushell sites. In other models the relationship between age and post-test score will be negative or very weakly positive.

Socio-economic Status

Strong

1. Within-model effects will favor high SES children in the Weikart model.

Weak--Based on 1969-70 Data

2. The overall interaction between SES and model will be insignificant for both the PSI and for the Stanford-Binet.
3. Within-model interaction effects on the Stanford-Binet will favor low SES children in Bank Street and Far West programs.

Ethnicity

Weak--Based on 1969-70 Data

1. Within-program interaction effects will favor white children in Engelmann-Becker and Tucson programs and Black children in Bushell programs.

Response Style

Weak--Based on 1969-70 Data

1. Within-model interactions will favor children making few competence responses in Engelmann-Becker programs. In Bank Street programs, interactions will favor those making many such competence responses. More generally, in more-directive programs interaction effects will favor children making few competence responses, while in less-directive models they will favor children making many such responses.
2. Within-model interactions will favor children making many passive responses in Bank Street, Bushell and Engelmann-Becker programs; in Weikart, Tucson and Far West programs interaction effects will favor those making few passive responses.

General Predictions

The introduction to this report raises three general questions about interactions. First, which characteristics of children interact most strongly with models--and, conversely, which are least important in terms of main effects and interactions? Second, what are the patterns of such interactions as do appear? Are the categories and

considerations which other researchers have used for classifying programs relevant to the interactions we observe? And third, how important are interactions in explaining cognitive outcomes of preschool programs? Neither the literature reviewed here nor the 1969-70 PV data provide final answers to these three questions, but they do suggest some hypotheses.

In answering the first question--that of the relative importance of different child variables in predicting interactions with model--it seems wise to consider not only the size and significance of interactions observed in the 1969-70 analysis, but also the degree to which these interactions are consistent across tests, and the interpretability of the patterns observed. Using these criteria I would make the following predictions:

1. Predictable, significant, interactions of model with ethnicity and SES will not be found.

2. For most models the effect of sex on test score will be small. For one or two models, however, the effect of this variable will be predictable and significant.

3. An interpretable interaction of age with model is unlikely.

4. Interactions of model with initial achievement, prior preschooling, and the two response style variables "passivity" and "competence", may follow patterns which are both predictable and interpretable.

The second general question raised in the introduction concerns the patterns of observed interactions and the relevance of categories which other researchers have used for grouping preschool programs to the interactions examined here. The patterns of interactions observed on 1969-70 PV data have been discussed in the preceding chapters. The grouping schema used by other researchers are helpful in interpreting

interactions of model with certain child variables. Most of these schemes relate in one way or another to the degree to which children's experiences and behavior are directed by the teaching adults. The interactions reported in the preceding pages suggest the following:

5. The "directiveness" continuum may apply to interactions of initial ability, prior preschooling, and "competence" with model; if an exception is made for Bank Street, this dimension may apply to interactions with "passivity".

6. The dimension of directiveness does not apply to interactions of model with sex. Interactions observed on the IQ measure may relate to quite another aspect of the learning environment: the degree to which adults use concrete objects in their teaching.

The third question we set out to answer is the most general: how important are interactions in explaining the cognitive outcomes of preschool programs? Can it be said that no one model produces maximum gains for all types of children? Here again our answer must be tentative, but the evidence of the 1969-70 data is that interactions of particular variables with model are quite important in explaining cognitive outcomes. For this analysis model effects are stronger on the PSI than on the IQ.[1] However, there is no one model or type of model which produces optimum gains for all types of children on either measure.

7. None of the eight models will produce optimum gains for all types of children across both cognitive measures.

---

[1]Model effects are significant for the PSI post-test analyses. They favor the achievement of children in Engelmann-Becker and Bushell models. Model effects on the Stanford-Binet are insignificant. Had the Fort Walton Beach site been included (See Appendix C) there would have been a significant model effect favoring Weikart.

PART II:

ANALYSIS OF THE 1970-71 PV DATA

Introduction to Part II

The hypotheses proposed at the end of Part I have been tested on the data generated in the second year of the HSPV study, 1970-71. Somewhat different methods of analysis have been used on the second year of data; the differences are described in the sections that follow.

The analysis of the 1970-71 data has two objects. The first is the testing of hypotheses. The second is a fuller exploration of the interrelations between the independent variables selected for study, and the relation between these variables and the eight models of preschooling. Part II, like Part I, is in seven sections, each one organized around the hypotheses proposed in Part I. These sections describe the analyses used, report the results, and summarize consistencies and inconsistencies between the two sets of data. Patterns suggested by the 1970-71 data are discussed.

The Sample

The sample for 1970-71 is similar to that used in 1969-70. Each of the eight sponsors[1] was responsible for implementing his model in three to five states. The sites were designated level I, II, or III. Planned Variations collected demographic information on children and staff in all sites; in levels II and III cognitive tests were also given. In level II sites these included the PSI but not the Stanford-Binet.

---

[1]Twelve sponsors were actually involved in this second year of Planned Variations. However, because of the strategy of hypothesis generation adopted in this report, only the eight models who were involved for both years have been considered.

In level III sites the PSI and other tests were given to all children. In addition the Stanford-Binet was given to a randomly selected half of each class.

In consequence, the sample of children for whom IQ data is available is considerably smaller than the sample of children with valid PSI scores. Specifically, there are 305 Black and white children[1] enrolled in the eight models for whom there are valid pre- and post-IQ and PSI scores. There are a total of 883 children in these eight models for whom PV has spring and fall PSI scores. In addition, there is a pool of control children; these children are enrolled in regular Head Start programs in the same (or comparable) communities as the various PV classrooms, but no sponsor is attempting to implement his model in their classrooms. For the eight models of preschooling examined in this report, there are 305 control children with valid scores on PSI and IQ pre- and post-tests.

The Analysis

The analysis of the 1970-71 data is directed at three general categories of question. First, what are the first- and second-order interactions of these child variables with the eight models of preschooling: do girls make greater IQ gains in Bushell while boys fare better in Far West; do children with prior preschooling out-score first-year children in Tucson and Bank Street, but not in Engelmann-Becker?

---

[1]Both Indian and Spanish-speaking children are excluded from these analyses. For other exclusions see Part I, Appendix C.

Second, what interactions not involving model significantly affect children's preschool achievement? Is it true, for example, that children of low SES show greater gains in the second year of preschooling than the first? Third, does grouping models as "more directive" and "less directive" add anything to our understanding of observed interactions between model and child variables? Do interaction effects become larger when several models are pooled together in a "model group", or do they wash out entirely, due to between-model differences?

The first set of questions--those involving interactions with model--have been investigated through analysis of covariance and regression.[1] Analyses of covariance are used to evaluate the magnitude and significance of model interactions with six categorial variables. These are sex, preschool experience, ethnicity, SES category,[2] and the two response style variables "passivity" and "competence". Regression analyses are used to investigate interactions with several continuous variables: initial IQ, initial PSI score,

---

[1]The 1970-71 analysis includes only ocvaried post-test scores. Raw gain scores are not analyzed separately in Part II because these interactions appear on the 1969-70 analysis to be quite similar to post-test interactions. Since the PSI and IQ tests are significantly correlated a multivariate analysis of covariance might have been done. However, the two tests are analyzed separately for two reason: first the PSI sample facilitates comparisons which would otherwise be impossible. Second, although the tests are correlated (.541 at pretest; see Appendix E) model effects on the two measures are quite different (see Smith, M.E., Some Short-term Effects of Project Head Start: A Preliminary Report on the Second Year of Planned Variation, 1970-71, Huron Institute, 1973). Although the reasons for observed differences are not always evident to the writer, it may be illuminating to the reader if the discrepancies are made apparent.

[2]For description of the SES measure, see Part I, Section 5.

age, and three SES components, mother education, income and family size. Appendix D gives further details on all analyses. Regressions are referred to by number in the text.

Effects and interactions not involving models are explored in a general interaction study analysis of covariance. The sample for this investigation is all children either in Planned Variations (sponsored) or comparison (unsponsored) classrooms with valid pre- and post-test scores on both the PSI and the Stanford-Binet. The primary purpose of this analysis is to test hypotheses concerning the achievement of children with prior preschool experience.

When discussing a number of preschool models one is tempted to group them in some way. In the introduction to Part I it is noted that although researchers have used a variety of principles in grouping models, many of the resulting schemes have a good deal in common. Models which assume that adults should take a very directive role for a good part of the teaching day are put in one group, while models which assume that children should generally have choices about what they do (and how) end up in another. Some models don't fit either description; in consequence, nearly all researchers use three or more categories.

In Section VIII of Part I, I suggested that concepts common to these traditional grouping schemes may be useful in predicting and interpreting interactions with particular child variables. In the 1970-71 analysis hypotheses relating to these groupings are investigated through regression analyses which are referred to in these pages as "grouped-model" regressions. These analyses are intended

to contrast models where adults take a highly directive role with models in which children make many significant choices about how--and with whom--to spend their time. Bushell and Engelmann-Becker are placed in a "more-directive" group, while Far West, EDC and Bank Street are placed in a "less-directive" one. In order to make the two groups as different as possible on the score of teaching strategy, the other three models, Weikart, Tucson and Gordon, are omitted from these analyses.

The groupings, and the choice of models to be included, are based on sponsor model descriptions. Bushell and Engelmann-Becker are placed in the more-directive category because in both models children's major learning is alleged to take place in adult-directed groups where decisions about what is taught are made by adults. Although adults in several other models are expected to plan small-group lessons (e.g., Tucson), the adult is generally expected to take account of the children's comments and demonstrated interests as he proceeds. Thus, the actual form of the lesson is supposed to be determined by children and teachers, working together (as in Weikart's "open" programs where both adults and children are expected to initiate and to respond [Weikart, 72]).

Far West, EDC and Bank Street are grouped together as less-directive because they emphasize the importance of the choices made by the individual child in his learning. All of these sponsors see the process of making these choices as essential to learning. None of them expect children to spend the major part of each day in an organized group, working with an adult.

Weikart and Tucson have been omitted because they are seen as falling somewhere between these two groups. Although these models emphasize the importance of choice-making, both appear to expect children to spend a significant part of each day in small groups with an adult. Although in practice either of these models might look quite a lot like the "less-directive" models, the structure for learning described by sponsors seems to direct the outcome of children's choices a bit more. The Gordon model is omitted because it contains no specific directions for classroom practice.

The Classroom Observation data supports the notion that practice in Engelmann-Becker and Bushell classrooms is quite different from that in other models (see Lukas and Wohlleb, 72). Engelmann-Becker and Bushell are significantly above all other models ($p < .001$) in total academic activity, and in frequency of adults with children in academic activities. They are below other models in independent child activity, with Bushell being significantly low ($p < .001$). Bank Street, Far West, Tucson, and EDC rank highest on this variable. In general the evidence for placing Engelmann-Becker and Bushell in the more-directive group is stronger than that binding Bank Street, EDC and Far West together. However, these three programs do show evidence of considerable independent child activity, both academic and non-academic.

I. INITIAL ABILITY

Section I of Part I utilizes three resources in formulating hypotheses concerning probable interactions of preschool model and initial ability: data presented in selected pre-school evaluation studies, Bar-Yam's review of certain aptitude treatment interaction studies, and the 1969-70 PV data. On the whole the investigation does not reveal strong patterns of achievement which are consistent across a number of studies and measures. It does, however, suggest three hypotheses:

Strong hypotheses

1. Less-directive models will show interactions favoring children of high initial IQ on the PSI post-test.

2. Weikart programs will show a stronger tendency than other programs to increase the variance of IQ scores.

Weak hypothesis

3. In Gordon programs within-model interaction effects on the PSI will favor children of low initial IQ. The variance of IQ scores will tend to decrease in this model.

The last of these predictions is derived solely from the 1969-70 PV data; the first two are supported by the findings of other investigators.

None of these three hypotheses is confirmed, as it stands, by the 1970-71 PV data. However, the analysis of this second-year data does suggest that the interaction of model and initial ability will be

along the lines of our hypothesis when the PSI rather than the IQ test is used as the measure of initial ability or achievement.

The 1969-70 analyses employ two strategies for exploring interactions of initial ability and model. First, the variance of pre- and post-test IQ scores are contrasted in each of the eight models. Second, children are stratified according to their initial IQ score and then the relative performance of "high" and "low" scoring groups are compared within each model.

The approach used in the 1970-71 analysis is different in two ways. First, children are grouped on the basis of PSI pre-score as well as IQ pre-score. Second, the initial ability measures are considered as continuous rather than categorical variables. To permit this more efficient use of the data, interactions are explored through regression rather than (as in the 1969-70 data analysis) analysis of covariance.

The 1970-71 PV Data: Results of Hypothesis Testing

Hypothesis 1: Less-directive models will show interactions favoring children of high initial IQ on the PSI post-test.

This hypothesis is based on the 1969-70 PV data and on results of certain ATI studies of older children. The first-year PV data show interactions favoring the PSI achievement of high IQ children in certain "less-directive" models (Bank Street, Tucson, and EDC). There is no substantial interaction effect for either of the "more-directive" models, Engelmann-Becker and Bushell. Bar-Yam, in a summary of ATI research, reports a similar pattern: in studies she reviews the effect

of initial ability level on achievement appears to be strongest in the more "permissive" programs.

In the 1970-71 analysis, the grouped-model regressions are used to test Hypothesis 1. In these analyses Engelmann-Becker and Bushell are classified as "more-directive" while EDC, Bank Street, and Far West are grouped together as "less-directive". The results of this analysis do not confirm the hypothesis as it stands. They do, however, suggest an interesting revision of it.

Independent variables in PSI regression 4a include model-group (as defined above), family size, mother's education, income, "competence",[1] "passivity", sex, ethnicity, preschooling, age, PSI pre-score, IQ pre-score, all first-order interactions involving model-group except model-group-by-PSI-pre-score, and selected second-order interactions (see Appendix D for complete list of independent variables). This analysis shows the model-group-by-IQ pre-score interaction to be significant ($p < .05$) but in the opposite direction from our prediction (see Table Ia).

In PSI regression 4b,[2] the variable "model-group-by-PSI-pre-score" is substituted for "model-group-by-IQ-pre-score". This variable is significant ($p < .001$) and in this case the effects are in the

---

[1] "Competence" and "passivity" are two child variables derived from the child's responses to the IQ test. See Part I, Section 7 for an explanation.

[2] Results of PSI regression 4b are used in preference to those for the larger sample PSI regression 3 because the distribution of initial scores for this sample poses a regression problem--children in one group pretest somewhat higher than those in the other group. There are no such between-group differences for the smaller sample.

Part II

Table Ia

Interaction of Model-Group With IQ Pre-Score
Effect on PSI Post-Test Score
(From PSI Regression 4a-Grouped Models)
1970-71 PV Data

Variable Description	Standardized Coefficient	Significance
Model-Group (less directive vs. more-directive)[1]	.019	NS
IQ Pre-Score	.330	.001
Group-by-IQ-Pre-Score	.106	.035

Effects on Adjusted PSI Post-Test Score[2]
(Given in Standard Deviations)

	Bank Street, EDC, Far West	Engelmann-Becker, Bushell
Low Initial IQ (One SD Below Mean)	-.243	-.417
High Initial IQ (One SD Above Mean)	+.205	+.455

[1]Dummy variable: Bank Street, EDC and Far West coded -1; Engelmann-Becker and Bushell coded +1.

[2]Effects given in the table combine main effects of the two variables and the interaction.

Part II

Table Ib

Interaction of Model Group With PSI Pre-Score
Effect on PSI Post-Test Score
(From PSI Regression 4b-Grouped Models)
1970-71 PV Data

Variable Description	Standardized Coefficient	Significance
Model-Group (less-directive vs. more-directive)[1]	.165	.025
PSI Pre-Score	.436	.001
Group-by-PSI-Pre-Score	-.215	.001

Effects on Adjusted PSI Post-Test Score[2]
(Given in Standard Deviations)

	Bank Street, EDC, Far West	Engelmann-Becker, Bushell
Low PSI pre- (One SD Below Mean)	-.816	-.056
High PSI pre- (One SD Above Mean)	+.486	+.386

[1]Dummy variable: Bank Street, EDC and Far West coded -1; Engelmann-Becker and Bushell coded +1.

[2]Effects given in table combine main effects of both variables and the interaction.

Part II

Table Ic*

Stanford-Binet and Preschool Inventory
Standard Deviations of Pre- and Post-Test Scores
Grouped by Model and Prior Preschool Experience
1970-71 PV Data

	Far West	Tucson	Bank St.	E.B.	Bushell	Weikart	Gordon	EDC
I. CHILDREN WITHOUT PREVIOUS PRESCHOOLING								
N	9	34	14	55	25	32	40	30
IQ Fall	9.130	12.207	11.430	12.727	10.994	11.547	14.210	11.309
IQ Spring	8.897	14.097	10.014	10.788	11.917	11.450	14.556	9.815
PSI Fall	13.039	10.576	10.315	10.496	8.035	9.836	11.612	10.893
PSI Spring	8.604	9.905	5.750	7.429	8.559	7.901	10.030	9.139
II. CHILDREN WITH PREVIOUS EXPERIENCE								
N	3	19	16	10	2	0	0	16
IQ Fall		7.711	8.451	14.160				10.813
IQ Spring		9.020	13.760	11.219				6.276
PSI Fall		8.905	10.142	9.932				10.153
PSI Spring		4.130	10.077	4.423				9.139

*Sample: all children with valid scores both on Stanford-Binet and on PSI, fall and spring.

predicted direction. The effect is more substantial and significant (see Table Ib) than that of model-group-by-IQ-pre-score.

The group-by-PSI-pre-score interaction is exactly as Bar-Yam had led us to anticipate: for children of high initial score the effect of curriculum (meaning model-group) on adjusted post-score is small, but for those of low initial score, curriculum has an effect of about three-quarters of a standard deviation, with children in the more-directive models scoring higher.

The difference between interactions observed for IQ and PSI pre-score probably relates to the fact that the PSI is not normed for age. High-scoring children, therefore, tend to be older as well as smarter. For this sample, the PSI pre-score is positively correlated with age ($r = .448$) and with IQ ($r = .404$). The IQ pre-score is, on the other hand, negatively related to age ($r = -.145$). Nonetheless, the observed interaction does not apparently reflect the effects of age alone: the model-group-by-age interaction is insignificant in both equations. Apparently, a measure which combines the effects of age and ability predicts achievement differences more powerfully than age or IQ alone.

Hypothesis 2: Weikart programs will show a stronger tendency than other programs to increase the variance in IQ scores.

Table Ic gives the variance in IQ and PSI scores, fall and spring, for all eight models. Variances are reported separately for children with and without prior preschooling. Clearly the hypothesis is not confirmed. Pre- and post-test variances do not differ significantly

for any model. However, for three models, Tucson, Bushell and Gordon, the variance in IQ scores does increase; it does not increase for the Weikart model.

> Hypothesis 3: In Gordon programs, within-model interaction effects on the PSI will favor children of low initial IQ. The variance of IQ scores will tend to decrease in this model.

This hypothesis is not confirmed. Table Ic shows that for children in Gordon programs the IQ variance increases (insignificantly) between fall and spring testing.

PSI regression 2 has been used to test the hypothesis concerning model-by-initial-IQ interaction in relation to PSI post-test score. Independent variables include models, income, mother's education, race, "competence"[1], age, PSI pre-score, IQ pre-score, interaction of model with mother's education, income, "competence", race, age, and IQ pre-score, and interactions of race and age with these variables. The model-by-initial-IQ interaction is significant ($p < .002$) for Engelmann-Becker; it is insignificant for all other models. For Engelmann-Becker the interaction favors children of low initial IQ.

Discussion

Three points emerge from the 1970-71 analyses. The first is that the structure and magnitude of an interaction of model with initial ability or achievement depends on what measure is initially used to

---

[1]Hertzig-Birch variable; see Part I, section VII for explanation.

categorize children. For these preschool analyses, interactions on the PSI are in one direction when children are classified on the basis of initial IQ and in the opposite direction when the PSI pre-test is used to group them. The greater magnitude and significance of the PSI-by-model-interaction suggest that a broad measure of initial achievement is more useful in predicting interactions than a measure which is adjusted for age.

These results fit well with data reported in Stodolsky's study of children's transition behavior (72). Stodolsky looked at the ways children move from one activity to another in a free choice situation. She reports correlations of age, mental age and IQ with frequency of each of ten behaviors (types of transitions, percentage of time spent in activity, etc.). Correlations of all observation variables with IQ are trivial, while correlations with mental age (Stanford-Binet score before age norming) are significant ($p < .05$) in six out of ten cases. (Correlations with age are even stronger than correlations with mental age; this may relate to the greater reliability of the age measure).

Stodolsky's results suggest that a measure which is not normed for age will give far more information about a child's actual classroom behavior than one which is. It is therefore not surprising that, in our analyses, PSI pre-score interacts more strongly (and more interpretably) with model-type than did initial IQ.

The second point emerging from these analyses is that the structure of an interaction may be quite different when the outcome measure is changed. The interactions of model-group and model with

IQ and PSI pre-test are significant only on PSI post-test analyses. On IQ analyses both interactions are insignificant. So, although these data indicate that more-directive models favor the PSI gains of initially low-scoring children, they give no evidence of a similar pattern on the IQ test.

These results are in line with findings of ATI studies (see Cronback, 1969; Bar Yam, 1969) that the structure of an interaction may often depend on the outcome sought. This fact should be borne in mind for all results reported in these pages. When an effect favors particular children in a particular model, the temptation is to say that these children "benefit" more. Yet often the advantage is specific to a particular test--we cannot assume that it extends to all cognitive measures, much less to the many other objectives of preschooling.

The third point is that the dimension of directiveness does appear, in a very limited and specific way, to affect interactions of initial achievement with model. The PSI grouped-model regressions show a substantial and highly significant interaction of PSI pre-test and group; this interaction is in the predicted direction.[1]

In sum, the 1970-71 PV data shows that when the PSI is used as a measure of initial achievement, children's performance in the PSI

---

[1]The strongest interactions of PSI pre-score with model are for EDC and Engelmann-Becker. PSI regressions Ib and Ic both show highly significant interactions ($p < .01$) for these two models, For EDC effects favor high-scoring children while for Engelmann-Becker the opposite is true.

post-test follows the patterns which the ATI literature had led us to predict, and that the interaction of achievement with model is substantial, particularly for initially low-scoring children. Beyond this, the analyses indicate that the interaction cannot be generalized to other cognitive tests and, in particular, does not apply to IQ. These findings are not strong, but they are of interest, particularly as they dictate caution in generalizing across measures.

II. PREVIOUS PRESCHOOL EXPERIENCE

Seven hypotheses relating to the effect of prior preschooling on children's achievement are proposed in Part I.

Strong Hypothesis

1. The cognitive gains of children having previous preschool experience will be smaller than those of children having no such experience.

Weak Hypotheses

2. Children with a particular type of preschool experience will make greater second-year IQ gains in a program which is different from their first experience than in one which is similar to it.

3. Children whose first preschool experience was in a highly-structured program will be more likely to maintain or increase IQ gains during the second preschool year than children whose first-year program was less-structured.

4. Children making large IQ gains in their first preschool year will make smaller gains during the second year than children making negligible gains the first year.

5. Children of low initial IQ will make greater gains during the second year of preschooling than children of high initial IQ.

6. Children of low SES will make greater second year IQ gains than those of high SES.

7. In Tucson and Bank Street programs, interaction effects on both cognitive measures will favor children with previous preschooling; in Gordon programs, interaction effects will favor children without such experience. In Engelmann-Becker programs, interaction effects on the PSI will favor those without previous preschool experience.

The first six hypotheses are based on data reported by other researchers; the first one is supported by the 1969-70 PV data. The last prediction is based on the 1969-70 data alone.

Three hypotheses (1, 6 and 7) have been tested on the 1970-71 data. The hypothesis which rests solely on the data of other investigators is not confirmed. Predictions based on the 1969-70 data are to some degree supported by the second year's analysis.

The results of the 1970-71 analyses suggest that we may be able to predict what models and what type of model will prove especially effective in raising scores of children with prior preschooling. The analyses do not indicate that certain types of children will gain more from a second year of preschooling regardless of the character of the program.

The 1970-71 PV Data: Results of Hypothesis Testing

Hypothesis 1: The cognitive gains of children having previous preschool experience will be smaller than those of children having no such experience.

The general interaction analyses of covariance have been used to test this hypothesis. Four analyses have been done, with IQ post-test, PSI post-test, IQ gain and PSI gain as dependent variables. The design is "competence"[1] by "passivity"[1] by preschool experience by ethnicity by SES. PSI pre-test, IQ pre-test, and age are covariates on the post-test analyses; age is a covariate on gains analyses.[2] The effect of preschool experience is significant ($p < .001$) on all four analyses. Differences in gains, adjusted for the covariates given above, amount to five points on the IQ (children with prior

---

[1] See Part I, Section 7 for explanation of these variables.

[2] For more details on analyses see Appendix D.

preschooling gain 1.1 points, while those without it gain 6.1 points). The difference in gains is four points on the PSI (8.5 points for those with preschooling versus 12.5 for those without). The hypothesis is confirmed.[1]

Hypotheses 2, 3, and 4:

2. Children with a particular type of preschool experience will make greater second year IQ gains in a program which is quite different from their first experience than in one which is similar to it.

3. Children whose first preschool experience was in a highly-structured program are more likely to maintain or increase IQ gains during the second preschool year than children whose first year program was less structured.

4. Children making large IQ gains in their first preschool year are less likely to make substantial gains during the second year than children making negligible gains the first year.

Planned Variation's information on the character of children's previous preschool experiences and on their cognitive gains prior to entry in the model is not adequate to permit testing of these three hypotheses.

Hypothesis 5: Children of low initial IQ will make greater gains during the second year of preschooling than children of high initial IQ.

In Part I, Section 2 the term "initial IQ" is used to refer to the child's IQ at the beginning of the first preschool year. For children with prior preschooling this information is not available

---

[1]Nevertheless, as is pointed out in Part I, the gains of second year children are not trivial when compared to the gains to be expected for such children were they not enrolled in preschool. See Smith, Marshall, Some Short-term Effects of Project Head Start: A Preliminary Report on the Second Year of Planned Variation--1970-71, Huron Institute, 1971.

to the Planned Variation Study. Indeed, most children had never been tested before their participation in Planned Variations. Therefore, the hypothesis as it stands could not be tested.

Hypothesis 6: Children of low SES will make greater second year IQ gains than children of high SES.

The general interaction analysis shows children of low SES making significantly greater IQ gains than those of high SES in both the first and the second preschool year. This effect seems, however, to result mainly from their lower prescores, since the effect of SES on covaried post-test score is insignificant in the general interaction analyses. The IQ analyses show no significant interaction either of SES or its components (family size, mother education, and income) with preschooling.

Hypothesis 7: In Tucson and Bank Street programs, interaction effects on both cognitive measures will favor children with previous preschooling; in Gordon programs interaction effects will favor children without such experience. In Engelmann-Becker programs, interaction effects on the PSI will favor those without previous preschool experience.

Table IIa gives the estimated combined means on the PSI post-test for children with and without prior preschooling. Interaction effects are significant for Tucson ($p = .006$), confirming the hypothesis for this model. For all other models interaction effects are insignificant.

Because the IQ sample is smaller and somewhat unbalanced with respect to prior preschooling, the hypothesis as stated here could not be tested on the Stanford-Binet data. It is, however, possible to test the broader form of the prediction which is given in

Part II

Table IIa

Interaction of Model With Prior Preschooling
Estimated Combined Means* for PSI Post-Test Scores
1970-71 PV Data
Grouped by Prior Preschool Experience and Model

	Far West	Tucson**	Bank St.	Engelmann-Becker	Bushell	Weikart	Gordon	EDC
Prior Preschooling								
Est. Mean	44.09	55.21	45.84	49.45	48.25	51.07		48.49
N	17	48	30	22	6	7	0	31
No Prior Preschooling								
Est. Mean	49.16	47.57	48.51	50.36	50.13	51.03	48.49	48.46
N	51	80	79	118	100	87	100	100

*Design: model by ethnicity by preschool experience. Covariables: age, PSI pre-score, SES, and sex.

**Effect favoring those with prior preschooling significant ($p < .006$).

Section 2, Part I:

> In a more generalized form [the hypothesis] might be stated "child-centered" programs will show interaction effects favoring second-year children while programs where the initiative lies primarily with adults will show interaction effects favoring first-year children.

This hypothesis is confirmed by the grouped-model regressions. In IQ regressions Ia and Ib the interaction of model-group with prior preschooling is significant ($p < .005$) and in the predicted direction. Table IIb shows children with no prior preschooling gaining somewhat more in more-directive models while second-year children appear to do substantially better in less-directive programs.

Discussion

Part I, Section 2 raises two questions about the effect of prior preschooling. First, are there some categories of children who, regardless of model, benefit more than others from a second year of preschool? Second, do some educational approaches work particularly well for children with prior preschool experience? Hypotheses relating to both questions are proposed. Specifically, it is suggested that children of low SES and low initial IQ might benefit more from a second preschool year; it is also suggested that children who had failed to make substantial gains in their first year of preschooling might be more likely to gain in the second year. Not all these hypotheses can be tested on the PV data. It has, however, been possible to investigate in considerable detail the possibility that some children gain more than others from a second preschool year, and the results of this investigation are essentially negative. The general interaction study reveals no significant interaction of prior

Part II

Table IIb

Interaction of Model-Group with Prior Preschooling
Effect on Stanford-Binet Post-test Score
(From IQ Regression 2a-Grouped Models)
1970-71 PV Data

Variable Description	Standardized Coefficient	Significance
Model-Group (more-directive vs. less-directive)[1]	-.123	.147
Preschooling[2]	.184	.006
Group-by-Preschooling	.203	.004

Effects on Adjusted IQ Post-Test Score[3]
(Given in Standard Deviations)

	Bank Street, EDC, Far West	Engelmann-Becker, Bushell
Prior Preschooling	+.142	-.510
No Prior Preschooling	+.104	+.264

[1]Dummy variable: Bank Street, EDC and Far West coded -1; Engelmann-Becker and Bushell coded +1.

[2]Dummy variable: children with prior preschooling coded -1; those without such experience coded +1.

[3]Effects given in the table combine main effects of both variables and the interaction.

preschooling with sex, age, ethnicity, SES, or either cognitive style measure. IQ regression 2 and PSI regression 4 reveal no significant interaction of previous preschooling with the three SES components, family size, mother's education and income. Although undoubtedly some children benefit more than others from a second year of preschool, there is no evidence that these children can be described in terms of the variables used in this report.

Although the 1970-71 analysis does not isolate one group which, regardless of program, benefits more than other groups from a second preschool year, it is somewhat more successful in locating models which are particularly beneficial to children with prior preschooling. Where significant interactions of preschooling with model or model-group (models grouped as "more-directive" and "less-directive") are found, they tend to support the hypothesis that more-directive models--or ones where, in Weikart's terms, adults initiate--will favor the achievement of first-year children, while less-directive models--those where the child initiates--will favor the achievement of second-year children.[1] Although the interaction of model-group with preschooling is not consistently significant in PSI analyses (PSI regressions 4a and 4b), it is significant ($p < .005$) for the IQ analyses. Table IIb indicates that for children with prior preschooling, model-group makes a substantial difference: children in the less-directive models outscore those in more-directive models

---

[1]The main effect of model-group ("more-directive" vs. "less-directive") on IQ post-test score is not significant.

by about two-thirds of a standard deviation. For children without preschool experience the effect is smaller but in the opposite direction, as predicted.

In Part I, I suggest that in models which require children to make choices and take initiative, children with earlier school experience might possess an advantage relative to those without it, while in more-directive models this might not be so. These data support that interpretation, although within-model effects in the PSI are not totally consistent across the two years of data.

III. SEX

Neither the preschool studies reviewed in Part I nor the 1969-70 PV data lead us to expect strong interactions of sex and model. Of the three hypotheses proposed in section III of Part I the two strongest--those based upon the findings of other researchers as well as on the 1969-70 PV data--are essentially negative in character.

Strong Hypotheses

1. There will be no consistent main effect of sex on cognitive outcome measures.

2. Within models the effect of sex on cognitive gains will differ from site to site.

Weak Hypothesis

3. In Tucson and Far West sites IQ differences will favor boys; in Bushell sites IQ differences will favor girls.

The patterns suggested by the 1969-70 data and by previous research on preschool achievement are in general repeated in the 1970-71 analyses: the effect of sex on test scores is for the most part small; where interactions of sex and model are predicted, the observed effects are in the predicted direction, but do not reach significance. However, analysis of the 1970-71 PSI data suggests that although the interaction of sex and model may be trivial when the sample is composed of children without prior school experience, it may be considerably more significant after the first preschool year.

The 1970-71 Data: Results of Hypothesis Testing

Hypothesis 1: There will be no consistent main effect of sex on post-score and gains for PSI and IQ.

This hypothesis is confirmed. The General Interaction Study analysis of covariance shows insignificant effects favoring boys on both the IQ post-test and the PSI. These effects are very small: .15 points on the Stanford-Binet (.01 SD's) and 1.36 points (.13 SD's) on the PSI. For the PSI sample (PV only), by contrast, regression analysis shows a small (.56 points, .05 SD's) but significant effect favoring girls on the PSI post-test. These effects are neither large nor consistent.

Hypothesis 2: Within models the effect of sex on cognitive gains will differ from site to site.

The effect of sex on adjusted PSI post-score is trivial in all models. Table IIIa gives the estimated combined means for boys and girls on the IQ post-test. These means are adjusted for IQ and PSI pre-score, age, preschool experience, SES and ethnicity. For five models (Tucson, Bank Street, Engelmann-Becker, Weikart, and EDC) Planned Variations has relevant IQ data for children in more than one site. Differences between the scores of boys and girls are trivial (less than one point) or favor boys in all Gordon, Engelmann-Becker and Tucson sites. In the two Bank Street sites differences favor girls. Sex effects are in opposite directions only in EDC sites.

Hypothesis 3: In Tucson and Far West sites IQ differences will favor boys; in Bushell sites IQ differences will favor girls.

Part II

Table IIIa

Estimated Combined Means for Stanford-Binet Post-Test Scores
Grouped by Model, Site and Sex
1970-71 PV Data

SITE	1	2	3	4	5	6	7	8	9	11	12	13	14
Boys Estimated Mean	91.10	99.12	95.10	88.12	89.46	98.25	93.04	93.73	115.4	100.6	91.42	91.10	95.85
N	7	10	20	4	10	11	18	7	16	11	11	4	13
Girls Estimated Mean	88.01	95.36	95.16	90.13	95.36	97.65	91.03	93.37	119.5	96.17	89.05	92.79	93.20
N	5	11	12	6	10	13	23	20	13	9	9	6	23
	Far West	Tucson		Bank Street		Engelmann-Becker		Bushell	Weikart	Gordon		EDC	

Design: model by ethnicity by "passives" by "competence" by SES by sex.

Covariates: IQ pre-test, PSI pre-test, age.

Table IIIa gives the estimated combined IQ means for boys and girls in each of the fourteen test sites. For all sites except site 3 sex effects are in the predicted direction; they are not, however, significant.

It is suggested in Part I that the tendency of certain models to favor the IQ achievement of boys—the Far West model in these PV analyses, Montessori in Bissell's analysis—may relate to an emphasis these programs place on learning through manipulation of concrete objects. Although interaction effects favoring boys are not statistically significant (due, probably, to the small number of children in Far West classrooms), they are in the direction predicted. This suggests that the possibility of such a program emphasis being particularly favorable to boys' IQ gains should perhaps be investigated by other researchers who have more classroom observation data available to them.

Discussion

In relation to sex the 1970-71 PV data follow the patterns observed in the 1969-70 data. First, main effects of sex are small and not consistent across the two cognitive tests. Second, sex effects within models are somewhat greater for the IQ than the PSI. And finally, sex effects on the Stanford-Binet are in the direction hypothesized for the three models for which predictions were made.

Although the model analyses indicate that we may be able to predict the direction of sex differences in a few models, they provide no evidence that these differences will reach levels of statistical

or educational significance. However, the 1970-71 grouped-model PSI regressions raise the suspicion that, in fact, differences between the achievement of boys and girls within a particular type of curriculum may be more substantial after the first preschool year. All PSI grouped-model regressions show a substantial and significant second-order interaction involving sex, model group, and preschooling. (The interaction is in the same direction on IQ analyses, but is far too small to reach statistical significance.) Table IIIb gives the magnitude and direction of effects. For children without prior preschooling the relationship between sex, model-group and achievement is basically additive, with girls achieving more than boys, and children in directive models achieving more than those in less-directive models. For children with pre-school experience, however, the situation is different. Although main effects continue to favor the achievement of girls and of children in more-directive programs, the sex-by-model-group interaction is strong: in Engelmann-Becker and Bushell, boys outscore girls by about one-fifth of a standard deviation, while in less-directive models, girls outperform boys by nearly one-half a standard deviation. Table IIIb suggests that curriculum assignment affects the achievement of boys with prior preschooling far more than that of first-year boys.

These results are not easy to explain, but they do raise some interesting questions. Why does model assignment have so little effect on girls' scores and so strong an effect on the achievement of second-year boys? Why do second-year boys outperform second-year girls

Part II

Table IIIb

Interaction of Model-Group, Sex and Prior Preschooling
Effect of PSI Post-Test
(From PSI Regression 3-Grouped Models)
1970-71 FV Data

Variable Description	Standardized Coefficient	Significance
Group[1]	.184	.001
Sex[2]	.069	.034
Preschooling[3]	.087	.020
Group-by-sex	-.073	.087
Group-by-sex-by-preschooling	.101	.014

Effects on Adjusted PSI Post-Test Score[4]
(Given in Standard Deviations)

Children with No Prior Preschooling

	Bank St. Far West EDC	Engelmann-Becker, Bushell
Boys	-.132	+.174
Girls	-.056	+.368

Children with Prior Preschooling

	Bank St. Far West EDC	Engelmann-Becker, Bushell
Boys	-.514	+.202
Girls	.028	-.008

[1]Dummy variable: Far West, EDC and Bank Street coded -1; Engelmann-Becker and Bushell coded +1.

[2]Boys coded -1; girls coded +1.

[3]No prior preschooling coded +1; Prior preschooling coded -1.

[4]Effects given in the table main effects of all variables and interactions.

in more-directive models? Will the differences observed for children with prior preschooling carry over into the school years?

IV. AGE

Although the studies reviewed in Section I show no interpretable pattern of age-by-program interaction, several researchers report results which indicate that within particular programs age may be strongly related to gains. The 1969-70 PV data is consistent with this: for children without prior preschooling the relationship between age and cognitive gains appears to be different in different models, and, within models, remarkably consistent across the two cognitive tests. Furthermore, the age-by-model interaction observed in this first year of PV data does not appear to relate to identifiable features of PV models.

On the basis of the 1969-70 PV data two weak hypotheses are proposed:

1. The relationship between age and adjusted post-score on PSI and IQ tests will be stronger and more significant for children without previous preschool experience than for children with it.

2. For children without previous preschooling, age will be positively related to IQ and PSI post-test in Tucson and Bushell sites. In other models the relationship between age and post-test score will be negative or very weakly positive.

Neither of these predictions is made with great confidence, since the author was unable to arrive at a plausible explanation of the observed interactions. Neither is confirmed by the 1970-71 analysis. For the 1970-71 PV data the pattern of age-by-model interactions is the same for children with and without prior preschooling. And the Bushell model, far from favoring older children, shows a weak effect favoring

younger children on the PSI.

The 1970-71 PV Data: Results of Hypothesis Testing

> Hypothesis 1: The relationship between age and adjusted post-test score on PSI and IQ tests will be stronger and more significant for children without previous preschool experience than for children with it.

This hypothesis is not confirmed. The interaction between age and preschool experience is not significant on any analysis of PSI or IQ post-score.

> Hypothesis 2: For children without previous preschooling, age will be positively related to IQ and PSI post-test in Tucson and Bushell sites. In other models the relationship between age and post-test score will be negative or very weakly positive.

The possibility of age-by-model interactions has been investigated through regression analysis.[1] These analyses, IQ regressions 3 and PSI regression 5, use only children without prior preschooling. Independent variables in IQ regression 3a include models, IQ and PSI prescore, sex, age, ethnicity, SES and response style variables, and selected interactions which are significant on previous analyses (see Appendix D). Interactions of age with the Tucson and Bushell models are forced in. Both are insignificant. In IQ regression 3b main effects are forced in and all age-by-model interactions are permitted to enter stepwise.

---

[1] In the 1969-70 analyses the relationship between age and covaried PSI and IQ post-test scores was investigated through regression equations. A separate equation was used for each model (see Part I, section IV for a list of other variables in the equations). The regression equations used for the 1970-71 analysis include children in all eight models; age-by-model interaction terms are used to evaluate the significance of differences between models.

None are significant.

Independent variables in the PSI regression include model, family size, mother's education, income, sex, race, age, PSI pre-score, and the two-way interactions of all these with model, age and ethnicity. The main effect of age is insignificant. The age-by-model interaction is significant ($p < .005$) only for Engelmann-Becker; it is negative as predicted. The age-by-model interaction is insignificant for Tucson and for Bushell.

Thus, the 1970-71 PV data does not support this hypothesis.

Discussion

On the whole, the 1970-71 PV data tend to support the prediction made earlier (section VIII, Part I) that an interpretable interaction of age with model is unlikely. These data do not confirm either hypothesis generated from the 1969-70 data. Furthermore, such interactions as are observed are not consistent across the two tests.

It is, however, of some interest that in the 1970-71 PSI data, in contrast to the 1969-70 data, the two Behaviorist models, Engelmann-Becker and Bushell, seem to fall together in favoring the achievement of young children. While these interactions are not significant across all analyses, effects are consistently in this direction. For the less-directive models the picture is more mixed; in consequence the grouped-model regressions show no significant interaction of age with model group.

V. SOCIO-ECONOMIC STATUS

The preschool evaluation literature provides ample basis for a hypothesis that the interaction of SES and model will favor low SES children in highly-directive models such as Engelmann-Becker and Bushell. The 1969-70 PV analyses, however, raise the suspicion that although this may be so in experimental preschool studies, it is not true when the setting is Head Start. On the basis of these data, three hypotheses are proposed:

Strong Hypothesis

1. Within-model effects will favor high SES children in the Weikart model.

Weak Hypotheses

2. The overall interaction between SES and model will be insignificant both for the PSI and for the Stanford-Binet.

3. Within-model interaction effects on the Stanford-Binet will favor low SES children in Bank Street and Far West programs.

The 1970-71 data support the notion that at least in the context of Planned Variations Head Start, more-directive models are not consistently more effective in raising poor children's test scores.

The SES measure used in the 1969-70 analysis combines family size, mother's educational achievement (years in school) and income, standardized and weighted equally. This measure is also used on the 1970-71 data.[1] Analysis of covariance done on PSI and IQ post-test

---

[1]The procedure used to determine each child's SES score is described in Part I, Section V. For the 1970-71 PV sample, the means are as follows: income, $4,350; mother's education, tenth grade; household size, 5.4.

scores gives no indication of an interaction between SES--so defined--and model.

Because other researchers (see Part I, Section V) have found a fairly consistent pattern of SES-by-program interactions, I have analyzed the 1970-71 PV data further, to see whether the components of our SES measure, taken singly, interact more strongly with model, and whether any observed interactions follow the patterns observed by Bissell (70) and others.

Specifically, I wanted to know whether, within the PV sample, more-directive models favor the achievement of low SES children, while less-directive models favor the achievement of high SES children. Even though this is not true for a combined SES measure, it might have been true when SES was defined as income, mother's education, or family size alone. Although the observed interactions of SES components with model are of some interest, they do not indicate that low SES children--defined by these measures--gain more in highly-directive models.

Preliminary regression analyses of PSI post-scores using main effects and model-by-variable interactions of the three SES components (PSI regression Ia) show the main effect of all three variables to be insignificant. Significant interactions with model are found for income and for mother education, but not for family size. Since the effect of family size appears to be trivial, a second analysis (PSI regression Ic) includes main effects and model interactions for the two remaining SES variables, sex, race, preschool experience, age, PSI prescore, and age-by-preschool interaction. (For a full list of independent variables, see Appendix D). This analysis shows

significant main effects both for mother's education ($p < .001$) and for income ($p < .005$) and significant interactions of both with model.

The direction of the observed effects sheds some light on the results obtained for the combined SES measure. Although the main effect of mother's education favors children with more educated mothers, the effect of income is in the opposite direction, with children from poorer families doing somewhat better than those from less poor families.[1] On a combined measure these two effects cancel one another out.

The pattern of model interactions on the PSI is equally difficult to interpret: Engelmann-Becker shows significant interactions favoring low income children ($p < .001$) and those whose mothers have achieved a higher educational level ($p < .05$).[2] Interactions of other models with SES components are insignificant.

---

[1]These results raise questions about the relationship between these two variables. The observed main effects might make sense if, due to SES requirements for participation in Head Start, mother's education and income were negatively correlated. (This might happen if, for example, the income ceiling were set lower for children of college educated parents than for those whose parents lacked high school diplomas.) Alternatively, they might be dismissed as the result of a suppressor effect if the intercorrelation was very high. In fact, however, the correlation is positive (.158), but too small to account for a strong suppressor effect. (See Appendix E for other intercorrelations in the 1970-71 PV sample.)

[2]There is some suppressor effect apparent here: the interaction of the Engelmann-Becker model with income becomes larger and more strongly significant when the interaction of mother's educational level with this model enters the equation. However, even when the effects of mother's education and income are considered quite independently (by examining the direction and magnitude of the partial correlations of each interaction term with the dependent variable on the step preceding entry of either interaction into the equation) interactions of the Engelmann-Becker model with these two SES components are in the opposite directions.

IQ regression Ic tests hypotheses relating to SES components. Independent variables in this equation include models, family size, mother's education, income, competence, passivity, sex, ethnicity, prior preschooling, age, IQ and PSI pre-score, and selected interactions of background variables which are significant in other analyses. Interactions of mother's education, income, and family size are allowed for those models for which interactions seemed likely. Specifically, with Weikart, Far West, and Bank Street, for which hypotheses had been proposed, and for Engelmann-Becker and Bushell, which Bissell's research identifies as likely to be different. Main effects are significant for family size ($p < .01$) and income ($p < .05$), favoring children from smaller families with higher incomes. The effect of mother's educational level is not significant. Only the interaction with family size is significant for Weikart, favoring smaller families.[1] When other interactions are allowed to enter the equation (IQ Regression Ia) effects are essentially the same except that interactions with Far West are replaced by a significant interaction of income with the Gordon model.

The 1970-71 Data: Results of Hypothesis Testing

Hypothesis 1: Within-model effects will favor high SES children in the Weikart model

Interactions of SES components with the Weikart model are not significant for the PSI post-test. However, for the IQ post-test,

---

[1] There is also an effect favoring children from smaller families in Far West. However, this interaction only reaches significance after the interaction of this model with income enters the equation (not significantly). Evidently the size of these correlated interactions is dependent on a suppressor effect.

the interaction with family size is significant ($p < .01$) and in the predicted direction: children from small families do better in this model.

Hypothesis 2: The overall interaction between SES and model will be insignificant both for the PSI and for the Stanford-Binet.

The analyses show this to be true when a combined SES measure is used.

Hypothesis 3: Within-model interaction effects on the Stanford-Binet will favor low SES children in Bank Street and Far West programs.

This hypothesis is not confirmed. Interactions of SES components with these two models are insignificant for both IQ and PSI analyses.

Discussion

These analyses indicate that a combined SES measure is not very useful for looking at effects of family background on performance in Planned Variation Head Start models. Perhaps this is partly because the SES sample represented in Head Start is truncated, consisting mainly of the very poor. Apparently, it is also because aspects of SES interact with the preschool models in different and sometimes contradictory ways.

On the whole, the interactions with model could not be said to follow the patterns which the results of previous research had led us to expect. The studies summarized in Part I, section V and, more fully, in Bissell (1970) would lead us to expect interactions favoring low SES children in Engelmann-Becker and Bushell programs and in the reverse direction for less-directive models. However, the

grouped-model regressions--PSI and IQ post-tests--show no significant interactions of model-group (more-directive vs. less-directive groups) with any of the three SES components. And although the PSI model regression does show a significant effect favoring low-income children in Engelmann-Becker, the effect is in the opposite direction for the mother-education measure, favoring children whose mothers have achieved a higher educational level.

It can be argued that all of the PV sample is poor and, in fact, comparable to the low SES group in some other studies.[1] According to this argument we might not expect SES-by-model interactions--the range of income, etc., being simply too narrow. However, for an all-poor sample such as this we would predict a strong main effect of model-group on both cognitive measures; we would expect the effect to favor more-directive models. The data does not support this expectation: when the most-directive and least-directive models are contrasted in the grouped-model regressions, the main effect of model-group on IQ post-test score is insignificant (favoring less-directive models). On PSI analyses the effect favors more-directive models, but its size and significance varies according to the other independent variables in the equation (compare, for example, Tables Ia and Ib in Part II; the effect of model-group is significant in one but not in the other).

---

[1] Although the SES data from Bissell's study is not quite comparable to the PV SES data, the indications are that the two samples are not very different. The median level of mother's education is between 9th and 11th grade for all three of Bissell's samples, as compared to 8th grade for 1969-70 PV sample and 10th grade for the 1970-71 sample. Median number of children in the family is 4 for Bissell's Urbana sample and 3 for the New York sample. This is comparable to the mean family size of 5.4 to 5.5 found for the two PV samples. Income data cannot be compared.

Why do the two years of PV data show a different pattern of results from Bissell's? In part this may be because the methods of analysis used are different. The PV analyses include more covariates and interactions than did Bissell's, and it is possible that the additional independent variables may pick up some variance shared with SES. This seems especially likely on the 1970-71 IQ analyses, since main effects for PSI pre-score and response style variables are included. However, this explanation is hardly a complete one, since none of these variables correlated above .131 with the three SES components (see Appendix E). Furthermore, these covariates are not included in the 1969-70 analyses; for the first round of data, covariates more nearly resemble Bissell's, with only prior preschooling and age added to the four she used (sex, ethnicity, prescore, and SES). There is no particular reason to suspect that prior preschooling or age relate to SES, although Appendix E reveals an unexpectedly high negative correlation between age and mother's educational level (-.197).

As Part I suggests, I would attribute differences between Bissell's results and those of the PV analysis to two differences between the two studies. First of all, the fact that the PV sample is national undoubtedly dilutes the strength of the SES measures. Second, the context of Head Start may change the character of the models in important and relevant ways. For example, pressure from parents and local communities may force teachers in particular sites to concentrate on the most disadvantaged children. The ideology of local bureaucracies and boards may influence teachers' responses to economic and social differences far more than the model does and hence may wash out SES-by-model

interactions. We cannot be sure whether this happens, but it is a possibility worth bearing in mind.

VI. ETHNICITY

The findings of previous preschool studies taken together with the 1969-70 PV data do not provide a basis for a strong hypothesis concerning ethnicity-by-model interactions. In the earlier studies reviewed the effect of race on cognitive gains is inconsistent. On the 1969-70 analyses, interactions of model with ethnicity are significant but exceedingly difficult to interpret. The two behaviorist models, for example, show opposite effects for ethnicity, with Bushell favoring Black children and Engelmann-Becker showing larger gains for white children. One weak hypothesis is proposed, based on the patterns of the first year's PV data:

> Within-program interaction effects will favor white children in Engelmann-Becker and Tucson programs and Black children in Bushell programs.

However, the puzzling pattern of observed interactions raises the suspicion that the observed interactions may have little to do with enduring characteristics of models and that although the interaction of ethnicity and model might continue to be of considerable magnitude, it is unlikely to follow predictable patterns.

This suggestion is supported by the 1970-71 data. Once again interactions of ethnicity and model are significant or (on the IQ) very close to it. And once again they follow no consistent pattern. The 1970-71 analyses support the notion that interactions of this variable with model do not relate to enduring characteristics of the models themselves.

The 1970-71 PV Data: Results of Hypothesis Testing

Hypothesis: Within-program interaction effects will favor white children in Engelmann-Becker and Tucson programs and Black children in Bushell programs.

This hypothesis is not confirmed. Interaction effects on the PSI post-test are significant ($p < .001$) for the Bank Street and Engelmann-Becker models, favoring white children in Bank Street and Black children in Engelmann-Becker (see Table VIa). The effect for Engelmann-Becker is in the opposite direction from that predicted on the basis of 1969-70 data. The interaction effect for Tucson is in the direction predicted, but does not reach statistical significance ($p < .06$). The interaction of ethnicity and model is not quite significant on the IQ analysis ($p < .06$). In Bushell and Engelmann-Becker models effects are in the predicted direction, but insignificant. The interaction of model and ethnicity is significant only in the Gordon model.

Discussion

The 1969-70 exploration of race-by-model interactions is limited at the start by the fact that only three models (Tucson, Bushell and EDC) have enough Black and white children to be included in both IQ and PSI analyses. The 1970-71 sample is better balanced, permitting inclusion of all models in both analyses. As it happens, the largest interaction effects in this round of data are for models excluded from the 1969-70 analyses. This, however, is only a partial explanation of our failure to predict the observed interactions. Even for models included in both years' analyses, there is little consistency

Part II

Table VIa

Estimated Combined Means* for PSI Post-Test Scores
Black and White Children
1969-70 PV Data

	Far West	Tucson**	Bank St.	Engelmann-Becker**	Bushell	Weikart	Gordon	EDC
Black								
Est. Mean	47.10	47.54	44.91	51.37	48.60	49.17	47.34	47.86
N	24	23	66	112	58	61	73	108
White								
Est. Mean	46.15	51.24	49.44	48.44	49.77	52.93	48.66	49.09
N	44	105	43	28	48	33	22	33

*Design: model by ethnicity by preschool experience.

**Effect significant, $p < .001$.

Part II

Table VIb

Estimated Combined Means* for IQ Post-Test Scores
Black and White Children
1970-71 PV Data

	Far West	Tucson	Bank St.	Engelmann-Becker	Bushell	Weikart	Gordon	EDC
Black Children								
Est. Mean	98.32	92.57	86.90	93.61	93.46	115.0	90.87	91.75
N	2	8	11	55	12	21	22	26
White Children								
Est. Mean	85.34	91.83	90.43	96.78	92.32	110.3	99.19	94.47
N	10	45	19	10	15	11	18	20

*Design: model by ethnicity by "passivity" by "competence" by SES by sex.

Covariates: IQ pre-score, PSI pre-score, age.

between the two years or between the two tests (compare Tables VIa and VIb in Part II, and Table VIa in Part I). Engelmann-Becker, for example, shows strong effects favoring white children in the 1969-70 PSI analysis, significant effects favoring Black children in the 1970-71 PSI analysis, and effects favoring white children in the 1970-71 analyses. Only Tucson shows a consistent pattern across the two years.

What can we then say about the interaction of ethnicity and model? This is still a puzzling question because, on the one hand, interactions of ethnicity with individual models continue to be statistically significant--more significant than interactions with many other variables--and, on the other hand, these interactions do not follow any consistent patterns.

In the 1970-71 analysis, ethnicity-by-model interaction effects for the PSI do appear to relate to the "directiveness" continuum discussed in the introduction. In 1970-71, three of the less-directive models, Tucson, Bank Street, and EDC, show effects favoring white children, while Engelmann-Becker, the more directive model, shows a highly significant PSI effect favoring Black children. However, no similar pattern is observed either for IQ scores or for the 1969-70 data. Furthermore, interactions of ethnicity with model-group (more-directive versus less-directive) are not significant for either measure, presumably because of opposite patterns shown by children in Engelmann-Becker and Bushell.

These puzzling differences between tests do not, as one might suspect, result from site effects. It was noted in the 1969-70 data that contradictory effects on the two tests for the EDC model

apparently resulted from one site, primarily white, scoring well on the IQ measure while children in the other site--mainly Black--did better on the PSI. However, confounding of race and site within models is not a problem for the 1970-71 sample. It is marked only for the Gordon sites, and this model happens to show consistent effects for the two tests.

We are left with the conclusion that although race-by-model interaction effects appear to be more important than the main effect of race in explaining cognitive outcomes,[1] these effects are neither interpretable nor predictable. The second set of data strengthens the suspicion aroused by the 1969-70 analyses, that observed interactions probably have little to do with identifiable characteristics either of ethnic groups or of models. They may result from idiosyncratic aspects of different sponsors' implementation strategies or from conditions within sites which relate only indirectly to models,[2] or from some other factor not considered.

---

[1]There is no consistent main effect of ethnicity on post-test score for the 1970-71 analyses. Although whites score slightly higher than Blacks on the IQ post-test, the two groups are not significantly different on the PSI post-test.

[2]Race effects are reasonably consistent across the sites within a particular model. That the observed interactions are not entirely due to specific site effects is demonstrated by the fact that the race-by-site interaction is insignificant for both outcome measures, while race-by-model interactions are significant, or very close to it.

VII. RESPONSE STYLE

The Hertzig-Birch scoring of the Stanford-Binet gives information about children's manner of meeting or avoiding the demands of the IQ test. In the 1969-70 analysis, this data is used to construct four "response style" variables. The variables are derived directly from four categories of responses: substitution, extension, competence, and passive. (For description of these categories, see Part I, section VII.) Children who, on the IQ pre-test, made more responses in a particular category than the sample median are described as "high" on that variable. Others are described as "low". Two of the variables, "competence" and "passivity"--named for the Hertzig-Birch categories on which they are based—show signs of interacting with model. This means that in some models children rated as "high" on "passivity" outscore those rated low on the variable, while for other models the opposite is true.

Two weak hypotheses, both derived empirically from the 1969-70 analyses, are formulated at the end of Part I, section VII:

1. Within-model interactions will favor children making few competence responses in Engelmann-Becker programs. In Bank Street programs, interactions will favor those making many such competence responses. More generally: in more-directive programs interaction effects will favor children making few competence responses, while in less-directive models they will favor children making many such responses.

2. Within-model interactions will favor children making many passive responses in Bank Street, Bushell and Engelmann-Becker programs; in Weikart, Tucson and Far West programs interaction effects will favor those making few passive responses.

These hypotheses have been tested on the 1970-71 data. The pattern of interactions is remarkably similar across the two years of data, suggesting that this type of variable may be quite useful in predicting interactions with model.

The 1970-71 PV Data: Results of Hypothesis Testing

> Hypothesis 1: Within-model interactions will favor children making few competence responses in Engelmann-Becker programs. In Bank Street programs, interactions will favor those making many competence responses. More generally: in more-directive programs interaction effects will favor children making few competence responses, while in less-directive models they will favor children making many such responses.

On the 1970-71 fall IQ test, most children made no competence responses at all. Therefore, children who made any such responses are categorized as "high" on this variable, while those making no such responses are classified as low.

The model-by-competence interaction is insignificant for both PSI and IQ model analyses. However, the grouped-model regressions do show a significant ($p < .005$) interaction in the predicted direction for the IQ post-test analysis. Table VIIa gives the magnitude and direction of main effects and interaction effects on the PSI post-test. The effect of curriculum assignment on post-test score is small for children who make no "competence" responses on the IQ pretest. It is sizeable, however, for those making "competence" responses: other things being equal, such children score more than half a standard deviation higher on the IQ post-test when assigned to "less-directive" models.

The group-by-"competence" interaction is not significant on the PSI post-test analyses, although it is in the predicted direction.

Part II

Table VIIa

Interaction of Model-Group with "Competence"[2] Effect
on Stanford-Binet Post-Test
(From IQ Regression 2a-Grouped Models)
1970-71 PV Data

Variable Description	Standardized Coefficient	Significance
Group[1]	-.123	ns
"Competence"[2]	-.097	ns
Group by competence	-.165	.004

Effects	Bank Street, EDC Far West	Engelmann-Becker, Bushell
Low in "competence" responses (0 responses)	+.055	+.139
High in "competence" responses (one or more responses)	+.191	-.385

[1]Dummary variable: Bank Street, EDC, and Far West coded -1; Engelmann-Becker and Bushell coded +1.

[2]Refers to Hertzig-Birch scoring of the Stanford-Binet. See Part I, section VII for explanation of this variable.

Hypothesis 2: Within-model interactions will favor children making many passive responses in Bank Street, Bushell, and Engelmann-Becker programs; in Weikart, Tucson and Far West programs, interaction effects will favor those making few passive responses.

This hypothesis is confirmed by IQ data. Table VIIb gives the estimated combined means for IQ post-test scores. The overall interaction is significant ($p < .04$), with effects in the predicted direction for all models. The interaction is not significant on the PSI post-test analysis.

Discussion

Interactions of the two Hertzig-Birch variables "competence" and "passivity" with model show quite similar patterns across the two years. Confirmation of the hypotheses relating to these variables is not overwhelmingly strong (no substantial interactions are observed on the PSI analyses). Nonetheless, the IQ analyses suggest that variables relating to cognitive style may be quite useful in predicting which children will make substantial gains within a particular model.

As observed in Part I, interactions of these variables with model seem to be related in some degree to the "directiveness" continuum. Broadly, we could say that more-directive models favor the achievement of children making some passive and/or no competence responses, while the less-directive models affect more positively the scores of children making some competence and/or no passive responses. The conspicuous exception to this formulation is Bank Street, which falls with EDC and Far West in favoring children high in competence responses, but is closer to the behaviorist model, Bushell, with respect to "passivity".

Part II

Table VIIb

Estimated Combined Means[1] for IQ Post-Test Scores
Planned Variation 1970-71
Children High and Low in "Passive"[2] Responses

	Far West	Tucson[3]	Bank St.	E-B	Bushell	Weikart	Gordon	EDC
Low in "passive" responses (0 responses)								
Estimated mean	94.24	97.95	85.58	95.17	90.79	114.8	95.74	94.06
N	10	46	19	35	18	16	29	30
High in "passive" responses (1 or more)								
Estimated mean	89.42	86.45	91.76	95.22	95.00	110.5	94.33	92.16
N	2	7	11	30	9	16	11	16

[1]Design: model by ethnicity by "passives" by "competence" by SES by sex.

[2]Covariates: IQ pre-test, PSI pre-test, age.

Refers to Hertzig-Birch scoring of Stanford-Binet. See Part I, section VII for explanation of this variable.

[3]Interaction effect significant, $p < .01$.

What is the reason for these patterns? Without direct observation in classrooms, it is impossible to be sure why a model produces more measurable growth for one group than for another. A few points are worth making, however. First of all, the evidence from preschool evaluations has suggested that very directive approaches favor the achievement of the educationally disadvantaged:[1] Bissell (1970) shows that low SES children do best in the most directive programs she examined; the analyses described in this report indicate that children without prior preschooling (section 2) and those initially scoring low on the PSI (section 1) achieve more in behaviorist models than in less-directive programs. The frequency of "competence" responses is, I think, one more aspect of "educational advantage", even though this variable is not highly correlated with SES, initial IQ, prior preschooling, or age. In section VII, Part I, it is suggested that children making many competence responses might have an advantage in less directive models because saying "I don't know" is a response likely to elicit help and information from adults--while passivity, refusal, or substitution may not. It is less clear why such children would be at a disadvantage in more directive models--unless adults in behaviorist models respond negatively to this style.

Taken together, the 1969-70 and 1970-71 analyses indicate that with respect to children making many passive responses, Bank Street teachers may act differently than teachers in other less-directive models. The model interactions with "passivity" suggest that in

---

[1] I use this term in its broadest sense to refer to groups of children who tend to do less well in school.

other less-directive models teachers may give less instruction and attention to children who respond passively to problems which they cannot solve unaided.[1] It may be that the Bank Street program is different because of its stronger emphasis on socio-emotional development. If the model focuses teachers' attention on goals in this area, it may make them more aware of the needs of children who act very withdrawn. Alternatively, socio-emotional growth may affect the IQ test scores of these children more directly than those of less "passive" children. Other interpretations are clearly possible; it would take direct observation in classrooms to be sure which aspects of the model are most relevant.

These two variables seem to be in some ways more interesting than the others considered in this report, because they related so much more directly to the child's actual behavior in a cognitive situation. We still do not know why a child behaves in particular ways during the IQ test. Nonetheless, the data suggest that model guidelines influence the way teachers respond to these different response styles. It seems likely (though as yet unproven), that if our index of response style were more sensitive we could see more substantial model effects. Perhaps we would then be able to judge more precisely which aspects of particular models are most salient in favoring the gains of one or another type of child.

---

[1]There is evidence that in at least one "open" preschool, teachers initiate more contacts with children who frequently seek them out. See Monaghan (1971) for documentation of the way in which patterns set by the children in the fall are maintained through the spring by the teachers.

PART III

DISCUSSION

Discussion

The introductory pages of this report raise three questions concerning the interaction of child and model characteristics. First, which characteristics of children interact most powerfully with model? Second, what are the patterns of interactions; are the considerations which other researchers have used to categorize programs relevant to the interactions we observe? And third, how important are interactions in explaining outcomes on cognitive tests? Section VIII of Part I presents tentative answers to these questions in the light of the 1969-70 Planned Variation data. Many of these predictions are supported by the 1970-71 analyses; some require further qualification.

General Predictions and Findings

1. Predictable, significant interactions of model with ethnicity and SES will not be found.

This prediction is supported by analysis of the second year's data. In the 1970-71 PV data, interactions of model with ethnicity are significant for the PSI. However, these interactions are not in the direction predicted on the basis of the 1969-70 PV data. For the combined SES measure--which weights equally family size, mother's educational level, and family income--interactions with model are trivial for both the PSI and the Stanford-Binet. Interactions of model with individual SES components are in some cases significant, but they do not follow a consistent pattern.

2. For most models the effect of sex on test score will be small. For one or two models, however, the effect will be predictable and significant.

Although sex differences on the Stanford-Binet are in the directions predicted, they do not reach statistical significance. For children without prior preschooling sex effects do appear to be small. There is, however, some indication that for children past the first year of preschool, sex may affect achievement more strongly: the PSI data show second year boys in more-directive models scoring substantially higher than boys in less-directive models; differences between the achievement of girls in more- and less-directive models are negligible.

3. An interpretable interaction of age with model is unlikely.

The 1970-71 data--particularly the IQ data--tends to support this prediction. However, there are some indications that on this round of data the two behaviorist models, Bushell and Engelmann-Becker, fall together in favoring the PSI achievement of younger children (this is not true in the 1969-70 analysis). In the less-directive models age appears to have inconsistent effects on test scores.

4. Interactions of model with initial achievement, prior preschooling, and the two response style variables "passivity" and "competence" may follow logical and predictable patterns.

This prediction is supported, in a limited sense, by the 1970-71 PV data. On the Stanford-Binet, interactions of the two response style variables with model and model-group are as hypothesized. Interactions on the PSI are not significant.

Interactions of model and model-group with prior preschooling also follow similar patterns across the two years of PV data: on both tests children with prior preschooling appear to gain somewhat more in the less-directive models, while first-year children do best in more-directive programs.

Interactions of initial achievement with model-group follow predicted patterns to some extent. However, this statement requires considerable qualification: the initial IQ variable, used in both 1969-70 and 1970-71 analyses, looks quite different on the two sets of data. But when initial achievement is measured by the PSI pre-score on the 1970-71 analysis, interactions--at least for the PSI--are as predicted. Children of low initial score do better in more-directive models than in less-directive ones, while for high-scoring children the opposite is true. However, because of the different patterns found for the two independent variables initial IQ and PSI pre-score, we must be very cautious in assigning importance to those results.

5. The "directiveness" continuum may apply to interactions of initial ability, prior preschooling and "competence" with model; if an exception is made for Bank Street, this dimension may apply to interactions with "passivity".

The 1970-71 PV analysis tends to support this prediction. As indicated above, the interaction of PSI pre-score with model-group is significant ($p < .001$), and in the predicted direction, for PSI grouped-model regressions. Interactions of prior preschooling and "competence" with model-group follow predicted patterns and reach acceptable levels of statistical significance ($p < .005$) on IQ analyses, although not on the PSI. Interactions of "passivity" with

model are as predicted on the IQ. Nevertheless, confirmation of predicted patterns is not as strong as it might be because in general the interactions reach significance on only one of the two tests.

6. The dimension of directiveness does not apply to interactions of model with sex. Interactions observed on the IQ measure may relate to quite another aspect of the learning environment: the degree to which adults use concrete objects in their teaching.

The 1970-71 data do not suggest striking confirmation of this prediction. Although sex effects within models are in the predicted direction for the Stanford-Binet, they do not even approach statistical significance. Furthermore, there is some indication that the dimension of directiveness may apply to interactions of model and sex in the PSI. The PSI grouped-model regressions show a substantial and significant second-order interaction of sex, model-group, and prior preschool experience. According to this analysis the contributions of sex and model-group to achievement are essentially additive for first-year children: girls do a little better than boys; more-directive models seem to boost scores a bit more than less-directive ones. For children with prior preschooling, however, the situation seems to be quite different. Although girls score about the same regardless of which type of program they are assigned to, boys do substantially (nearly three-quarters of a standard deviation on the PSI post-test) better in the more-directive models.

7. None of the eight PV models will produce optimum gains for all types of children across both cognitive measures.

This hypothesis is supported by the 1970-71 data. Although the main effect of the Weikart model on the Stanford-Binet is so strong as to dwarf interaction effects, this is not true for the PSI (see Smith, 73, for a detailed analysis of model effects on the PV cognitive battery). In terms of the IQ test, then, all types of children make largest gains in the Weikart model. However on the PSI the situation is different: for this test interaction effects tend to be as large as model effects, so that the model which produces optimum gains for one type of child is not the one which works best for another.

Summary and Conclusions

These general hypotheses provide a good starting point for a discussion of broad questions about the relative significance of different child variables, the usefulness of conventional model groupings in predicting interactions, and the overall importance of interactions in explaining cognitive outcomes.

The first question raised in the introduction concerns the relative importance of interactions of model with different variables or types of variables. As I noted in the discussion which concludes Part I, this is not always easy to evaluate, since the size and significance of an interaction does not necessarily tell us how important it is. Thus, ethnicity-by-model interactions are significant, or nearly so, for all analyses. But because the direction of the observed effects is consistent neither across years nor across tests this interaction does not help us to predict which children will benefit

most from particular models. In order to decide which variables are most important, it is necessary to consider the size and significance of effects, the degree to which they are consistent across two years of data, and the interpretability of the pattern observed. Using these criteria, the two response style variables, "competence" and "passivity" seem to be among the most important. Although interactions of these two variables with model and model group are insignificant on 1970-71 PSI analyses, the interactions are significant and in the predicted directions for the IQ analyses. On this measure, the pattern of observed effects is the same for both years of data, and makes sense in the light of what we know about preschool curricula.

To a more limited extent, we can say the same thing about prior preschool experience. Although not all the predicted interactions reach statistical significance, the pattern of effects is very similar across the two years of data: children with prior preschooling gain most in less-directive models, while first year children do best in more-directive programs.

For three other child characteristics, initial achievement, sex, and age, the 1970-71 analysis indicates the existence of interpretable interactions of some magnitude. However, since these interactions are unreplicated, I cannot speak about them with equal confidence.

Interactions of initial achievement, sex, and age with model appear to affect children's performance on the PSI post-test but not on the Stanford-Binet. The 1970-71 data suggest that behaviorist models favor the PSI achievement of young children, those initially scoring low in the test, and of boys with prior preschool experience.

Within-model differences are smaller and somewhat less consistent for less-directive models. The analyses relating to age and PSI pre-score lead this writer to infer that behaviorist approaches are especially efficient in facilitating the learning of children who have not yet reached a certain basal level of achievement; they suggest that other, more open-ended approaches may work better for children who because of age, prior preschooling, or natural precocity start the year somewhat better prepared.

These results parallel those reported by Bissell (70), in her analysis of the contribution of SES to final test score in different types of preschool programs, and to Bar-Yam's summary of ATI studies. Bissell's analysis shows that although children of higher SES outscore low SES children in less-structured or directive preschools, this is less often true in more-directive programs like Bereiter-Engelmann. Similarly, Bar-Yam reports that in a number of studies of older children's learning, students of low ability appear to gain more in "directive" programs than in "permissive" ones; for high-ability children the choice of curriculum appears to influence performance less. As noted earlier, these studies differ from Planned Variations in a number of ways. Nonetheless, the similarity of the patterns is suggestive.

Interactions of model and model-group with two remaining variables, SES and ethnicity, appear to be of a good deal less interest. Although interactions of model with ethnicity are significant in both 1969-70 and 1970-71 analyses, the patterns are inconsistent, even contradictory, across the two tests and the two years of data. Interactions of SES

and model are not significant in either years' analyses when SES is defined by a combined measure.

The PV data indicate that the finding of Bissell and others that more-directive models favor the achievement of low SES children may not hold when the models are implemented in the context of Head Start. Interactions of model-group (Engelmann-Becker and Bushell vs. EDC, Bank Street, and Far West) and SES components do not even approach significance either on IQ or PSI analyses. Furthermore, the main effect of model-group is insignificant (and favorable to less-directive models) on IQ analyses, and rather fragile on PSI analyses. (Although the effect favoring more-directive models is significant on some analyses, it is insignificant on others; the magnitude of the effect depends on the choice of other independent variables). Knowing that the whole PV sample is of low SES (compared to national norms), we had expected to find strong effects favoring more-directive models on both cognitive measures, but this is not the case.[1]

There are, I think, two general points we can make about the relative importance of the child variables considered in this report. First of all, the variables whose interaction with model follow the most consistent pattern across the two years of data are those which relate most directly to the child's behavior and experience as a learner: the response style variables, and prior preschool experience. Second, none of the variables which interact interestingly with model

---

[1] On the NYU booklet 4a, which tests knowledge of letters and numbers, the more-directive models do show stronger gains. This is not surprising, as the other models do not place major emphasis on this type of learning. (See Smith, 73.)

or model-group describe immutable characteristics of children. All of them describe the child at a particular point in his educational experience. Age, PSI achievement, prior preschooling, and response style: all these things change from year to year. Sex by itself shows no consistent interactions with model-group; only when it is considered in combination with prior preschooling is the interaction strong. The impact of particular models on little boys may depend on whether or not they have been in school before. IQ, which changes less over the years than achievement level, interacts far less powerfully with model and model-group.

This second finding--that the characteristics which interact most strongly with program type are those which are not immutable--is strikingly consistent with results reported in Stodolsky's observational study of children's choice-making behavior in several preschools (72). Stodolsky presents correlations of ten observational variables with age, mental age[1] (as measured by the Stanford-Binet), and IQ; she also reports effects of sex and SES on frequency of the observed behaviors. Correlations of observation variables with age and mental age are strong; sex effects on observed behaviors are at a minimum, but differences are significant for two of the ten variables; effects of initial IQ or SES on the ten observation variables are all insignificant. All Stodolsky's observational variables relate to children's behavior in preschool settings where they must decide for themselves

---

[1]Mental age (MA) as measured by the Stanford-Binet correlates .75 with PSI pre-score in the PV sample (fall 1970; see The Quality of the Head Start Data, 1973). The correlation of IQ with PSI score is lower (.54) because the IQ, unlike MA or PSI score, is standardized for age.

how to spend time; differences on such measures might well relate to children's achievement in less-directive models.

To me, these results mean one thing: the strategy which works best for a child today is not necessarily the one which will be optimum next month or next year. This is nothing new: plenty of good teachers use this knowledge every day in their classroom, allowing first-graders more freedom, for example, as their reading skills improve and they are more able to work independently. Nonetheless, the point needs emphasizing: all of the PV analyses described here lend weight to the idea that we can increase preschoolers' achievement by adapting curriculum in particular ways for particular children. None of the data support the notion that the choice of curriculum for a particular child or group of children should be final. Those characteristics of children which do not change--ethnicity, SES and sex--are precisely the ones which do not show consistent or interpretable interactions with model.

The second question which this investigation has sought to answer concerns the usefulness of various model groupings in predicting and interpreting interactions. In the discussion which followed Part I, I answered this question tentatively, saying that the 1969-70 data suggests that the dimension of "directiveness" applies to interactions of model with initial IQ, prior preschooling, and the two response style variables, but not to interactions of model with sex, ethnicity and age. The 1969-70 analyses raised the possibility that interactions of sex and model might relate to another dimension: the degree to which concrete objects are used for teaching and learning.

The 1970-71 analysis supports the idea that the dimension of directiveness applies to interactions of initial ability, prior pre-schooling and response style variables with model. But this second year's data suggests that the dimension may also relate, in limited ways, to interactions of age and sex with model. These interactions are observed only on PSI analyses, and not on the IQ. Sex effects on the IQ give very limited support to the theory that an emphasis on learning through concrete objects will favor the IQ gains of boys more than those of girls. However, the observed effects are not significant.

The dimension of directiveness thus looks more important to interactions observed in the 1970-71 analyses than to those found in 1969-70. But while saying this I want to emphasize again that in these analyses, the main effect of "directiveness" is small. Although children in directive models score a little bit higher on the PSI they do slightly (insignificantly) less well on the Stanford-Binet.

These observed differences between the two tests are consistent with early results reported by Robert Soar in his analysis of process and outcome in selected Follow-Through models (71). In the first two years' data, Soar found

> a tendency for abstract measures of pupil growth to relate positively to classroom behavior dimensions that reflect pupil freedom and self-direction, whereas simpler, more concrete measures of pupil growth tend not to relate, or even in some cases to relate negatively. In contrast, but relatively consistently, the simpler measures of pupil growth tend to be related to classroom behavior dimensions representing more structure and more control on the teacher's part.

These relations do not hold in Soar's analysis of later Follow-Through data (Soar, 1972). Nonetheless the early patterns are of interest because they parallel those reported for the 1970-71 PV analyses: on the PSI, a test heavily loaded with informational items, children in more-directive models score a little higher than those in the less-directive group. On the Stanford-Binet, by contrast, the very modest (and insignificant) differences between model-groups favor the less-directive models.

The third question we have asked about interactions of child characteristics and model concerns their overall importance. Can we say that one model or type of model is "best"--in the limited sense of maximizing cognitive gains--for all children? Or are interaction effects in fact more substantial than model effects? The 1969-70 analyses indicated that the answers to these questions were different for different tests; although on the Stanford-Binet interaction effects were more substantial than model effects, this was less often true for the PSI. For all models except Weikart, a similar pattern is observed in the 1970-71 analysis. The effect of the Weikart model[1] on IQ post-test scores is, however, so substantial as to dwarf the importance of interactions. (For more on this, see Smith, 1973).

On those of the 1970-71 analyses in which five models are grouped as more- or less-directive, the interaction of model-group with child characteristics explains substantially more of the variance in post-test

---

[1]Had the Fort Walton Beach Weikart site not been excluded in 1969-70 analyses the pattern across the two years would have been more similar.

scores--both PSI and IQ--than does the main effect of model-group. The main effect of model-group explains 1.9% of the unique variance in PSI post-test scores[1] and .5% of the variance in IQ scores.[2] Interactions of background variables with model-group explain more than 12% of unique variance in both measures. Tables in Part II which refer to grouped-model analyses (Ib, IIb, IIIb, and VIIb) indicate that neither approach (more-directive vs. less-directive) is optimum for all children.

This is less clearly true on the analyses where effects of the eight models are considered separately. Effects of models explain somewhat more variance, and interactions of model with child variables explain less. Nonetheless, for the PSI we can still say that no one model produces optimum results for all children (see for example, Table IIa, which indicates that Tucson maximizes PSI gains for children with prior preschooling, while Weikart favors those without preschool experience).

The model analyses of IQ post-test scores reveal quite a different situation. Although interactions of model with several child variables are significant, the main effect of the Weikart model on IQ post-test scores is so substantial as to reduce the importance of interactions. For 1970-71, we could say that, in terms of IQ, <u>all</u> types of children gain more in the Weikart model.[3]

---

[1]From PSI regression 3-grouped models.

[2]From IQ regression 2a-grouped models.

[3]Between-model differences are small for the other seven models, so that if Weikart were excluded, no one of the remaining models would produce optimum IQ gains for all types of children.

These findings have, I think, certain implications for educational policy and research. First, and most important, they support the idea that educational diversity can benefit children. Although the PV data indicate that one model, Weikart, may be astonishingly successful in promoting IQ gains with all kinds of children, they do not suggest that one type of model (more-directive vs. less-directive) maximizes cognitive gains for all kinds of children. The inconsistent patterns found in 1970-71 analyses on the PSI and IQ post-tests suggest that the choice of educational program should depend on the outcome sought as well as on the children, since main effects as well as interactions are somewhat different on the two tests. The Weikart model, for example, although outstanding in its effect on IQ post-test scores, is somewhat less effective than Engelmann-Becker in raising PSI scores.

If the PV analyses indicate that a choice of curriculum which takes into account the differences among children may raise test scores significantly, they also suggest that diversity should be created on the micro rather than the macro level--within schools, pre-school centers, or classrooms, rather than just within cities or school systems. I say this for two reasons. First, all the evidence from the PV analyses points to the fact that global demographic variables like ethnicity and SES do not interact in a predictable way with model, at least in a Head Start setting. Second, these data indicate that children's educational needs change--that while one model may efficiently raise the scores of four-year-olds without prior preschooling, another approach may benefit these children a year later. If, as I have argued, the variables which matter most are

those which relate to classroom behavior and learning style, then the design of curricula should be flexible. Otherwise half the benefits of diversity will be lost.

We have, I think, always known that no one educational approach works well for all teachers. If we can also demonstrate that no one type of curriculum is best for all children, then perhaps reformers should stop trying to change teachers' styles and instead start helping individual teachers to do what they see as a good job in the way in which they feel most effective. That effort, plus assignment of children which takes account of pupil needs and teachers' styles, might raise children's scores as well as teachers' morale.

What, if any, is the educational significance of the effects described in these pages? Given what we have begun to suspect both about the limited impact of school differences generally (Jencks et al., 72), and about the mortality of preschool IQ gains (Stearns, 71), it seems quite possible that a difference of half a standard deviation on the PSI or Stanford-Binet will not in future years translate itself into higher earnings, greater social mobility, or even improved understanding of fifth grade arithmetic. Nonetheless, differences of this magnitude do suggest that in the short-run, over the course of the preschool year, children are learning some kinds of things considerably faster than they were before the Head Start experience. The analyses described here suggest that for particular types of children some educational environments facilitate this learning more than others. And while these differences may make little impact in the long-run of people's lives, they may reflect some

important differences in the match between children's present needs and their preschool experience.

Given the very real limitations of the cognitive tests used here, and the great importance of other goals of preschooling, both cognitive and non-cognitive, we cannot be very sure that optimizing gains on the PSI or the IQ is of primary importance. But I do think it worthwhile to investigate why some children gain more than others in particular environments. The observational studies which would illuminate this point might well shed light on the first question: what kinds of growth do these test gains reflect, and how important are they anyway?

This report demonstrates, I think, the need for further research on the interaction of child and model variables. The specific findings discussed in these pages are nowhere near clear-cut enough to be confidently translated into classroom practice. Replication of any patterns reported here would be interesting and important. Nevertheless, one point seems to stand out: research directed at the question of what kinds of programs will benefit particular children in particular ways should look at characteristics of children which relate as directly as possible to their behavior in cognitive situations. This is no new idea: a number of good studies have done exactly this, often with interesting results (for a summary and review of several such studies, see Lesser, 1971). However, too little work of this sort has been done on the preschool level: we need more sensitive indices of response style and ideas about what other variables relate to children's classroom needs; we also need observation studies which

illuminate the reasons for observed interactions.

In designing studies which might help us understand which children are likely to make what kinds of gains in particular environments, we should bear two points in mind, both lessons of the Planned Variations Study. First, almost all classrooms provide a mixture of more and less-directive situations. While this may make interpretation of data more difficult, it provides a real opportunity for those interested in children's learning to observe one child in a range of learning situations and learn what "response style" means in practice. Perhaps we need instruments which help us observe the differences--and similarities--in ways in which children respond to more and less formal situations within the same classroom and ways children affect learning environments.

The second lesson which emerges from the analysis of Planned Variation data which is described here is that in order to learn about interactions we need small experimental studies designed to test specific hypotheses. In data dredging operations of the sort described in this report one too often lacks, in the end, the very information one needs in order to understand the most provocative findings.

APPENDICES

APPENDIX A

CORRELATIONS AMONG DEPENDENT AND INDEPENDENT VARIABLES; PLANNED VARIATIONS 1969-70

	age	sex	prior preschool	ethnicity	mother's ed.	income	family size	IQ pre-score	IQ post-score	PSI pre-score	PSI post-score	"extensions"	"substitutions"	"competence"	"passives"
age	--	.068	-.288	-.202	-.058	-.037	.078	-.157	-.203	.536	.498	-.103	-.231	.095	-.051
sex[1]		--	-.104	.115	-.019	-.018	-.035	.140	.073	.165	.161	.029	-.054	.068	.147
prior pre-schooling[2]			--	-.050	.153	-.098	-.025	-.118	.015	-.402	-.263	-.014	.042	-.063	-.112
ethnicity[3]				--	.075	-.018	.014	-.281	-.289	-.222	-.305	-.116	-.103	-.120	.002
mother's ed.					--	.228	-.018	-.027	-.020	.023	-.100	-.124	-.086	-.096	.032
income						--	.234	.093	.124	.107	-.005	.019	-.028	-.056	.000
family size							--	-.086	-.138	-.064	.016	-.117	.016	-.045	.090
IQ pre-score								--	.711	.394	.431	.138	-.082	.166	-.069
IQ post-score									--	.329	.383	.119	.015	.128	-.038
PSI pre-score										--	.644	.047	-.128	.093	-.027
PSI post-score											--	.067	-.153	.190	-.071
"extensions"[4]												--	.213	.063	-.009
"substitutions"[4]													--	.043	.007
"competence"[4]														--	-.144
"passives"[4]															--

[1]Dummy variable: boys coded +1; girls coded -1.

[2]Dummy variable: no prior preschool coded +1; prior preschool coded -1.

[3]Dummy variable: whites coded +1; blacks coded -1.

[4]Variables derived from Hertzig-Birch coding of Stanford-Binet pre-test. See Part I, section VII for a description.

APPENDIX A

SAMPLE SIZE FOR CORRELATIONS
PLANNED VARIATIONS 1969-70

	age	sex	prior preschool	ethnicity	mother's ed.	income	family size	IQ pre-score	IQ post-score	PSI pre-score	PSI post-score	"extensions"	"substitutions"	"competence"	"passives"
age	315	315	309	315	232	276	305	315	315	297	304	315	315	315	315
sex		--	309	315	232	276	305	315	315	297	304	315	315	315	315
prior pre-schooling			--	309	230	270	309	309	309	291	298	309	309	309	309
ethnicity				--	232	276	305	315	315	297	304	315	315	315	315
mother's ed.					--	232	223	232	232	219	223	232	232	232	232
income						--	267	276	276	260	265	276	276	276	276
family size							--	305	305	287	294	305	305	305	305
IQ pre-score								--	315	297	304	315	315	315	315
IQ post-score									--	297	304	315	315	315	315
PSI pre-score										--	286	297	297	297	297
PSI post-score											--	304	304	304	304
"extensions"												--	315	315	315
"substitutions"													--	315	315
"competence"														--	315
"passives"															--

APPENDIX B

PROGRAM EFFECTS ON HERTZIG-BIRCH RESPONSES AT TIME OF PRETEST

Because the Stanford-Binet pretest is given approximately three weeks after the opening of school, it is quite possible that different model emphases might affect children's style of response even on the initial test. It is of interest to know whether this is so--to know, in the terms of this thesis, whether, for example, a high proportion of children in one model score "high" (above the median) in passive responses, while most of those in another model score low. And if between-model differences of this sort exist, we must ask how they relate to the interactions reported in these pages.

Table A-1 shows the number of children in each model who score high and low in passive and competence responses for 1969-70 and for 1970-71. A χ^2 test has been used to evaluate the significance of between-model differences for each variable and each year. Differences are significant ($p < .01$) both years for incidence of passive responses, and, in 1970-71, for incidence of competence responses. Differences are not significant above the .05 level for incidence of competence responses for 1969-70.

What causes the observed differences and how do they relate to interactions reported in the test? If the differences derive from model effects on response style we would expect to find consistency from year to year. The patterns for the two years of data are not, in fact, strikingly consistent. This is especially true for competence responses: in 1969-70, 67% of the children in Bank Street

APPENDIX B

Table A-1

Number of Children Scored High and Low on Two Hertzig-Birch Variables
Planned Variations, 1969-70
Planned Variations, 1970-71

Table A-1a
Number of Children Scored High and Low on Competence Responses
Planned Variations, 1969-70

	Far West	Tucson	Bank St.	Engelmann-Becker	Bushell	Weikart	Gordon	EDC	Total
Low	8	25	33	32	8	21	19	21	167
High	8	22	11	29	6	22	13	26	137

Table A-1b
Number of Children Scored High and Low on Competence Responses
Planned Variations, 1970-71

	Far West	Tucson	Bank St.	Engelmann-Becker	Bushell	Weikart	Gordon	EDC	Total
Low	6	38	18	56	34	20	27	23	205
High	6	15	12	9	22	12	13	23	100

Table A-1c
Number of Children Scored High and Low on Passive Responses
Planned Variations, 1969-70

	Far West	Tucson	Bank St.	Engelmann-Becker	Bushell	Weikart	Gordon	EDC	Total
Low	6	28	20	23	5	24	25	27	158
High	10	19	24	38	9	19	7	20	146

Table A-1d
Number of Children Scored High and Low on Passive Responses
Planned Variations, 1970-71

	Far West	Tucson	Bank St.	Engelmann-Becker	Bushell	Weikart	Gordon	EDC	Total
Low	10	46	19	35	18	16	29	30	203
High	2	7	11	30	9	16	11	16	102

classrooms are classified as low in passive responses. In other models fewer than 60% of the children are scored low. In 1970-71, by contrast, somewhat fewer children in Bank Street are classed as low in competence responses than in the sample as a whole. The χ^2 value is attributable instead to the large number of Engelmann-Becker children scoring low on this variable, and the large number of EDC children scoring high.

The distribution pattern for passive responses is only slightly more consistent across the two sets of data. In 1969-70 the significance of between-model differences is mainly attributable to the large number of Gordon children making no passive responses, while in 1970-71 sizeable deviations from the general pattern are evident for Engelmann-Becker, Tucson, Bushell, and Weikart. The single consistency we observe is the large number of Engelmann-Becker children making passive responses both years.

Given the lack of consistency between 1969-70 and 1970-71, I would attribute most between-model differences to tester effects. I would make a tentative exception for Engelmann-Becker, saying that the data does suggest that even after a few weeks in that model, children may make more passive responses to IQ test items than we would otherwise expect. In view of the model emphasis on correct answers, this does not seem surprising.

How do the between-model differences reported here affect the interpretation of interactions between response style variables and PV models? First, the significance of between-model differences suggests that the Hertzig-Birch variables may not be as reliable as

we would hope--that they are probably affected to some degree by differences between testers. This is not surprising, but it is unfortunate. Second, the general absence of consistent model effects on pretest data suggest that children's initial response style is in most cases independent of model. A possible exception, Engelmann-Becker, may increase children's tendency to respond passively even at the time of pretest. Given the fact that this model may encourage this type of behavior it is somewhat surprising that the model does not show stronger effects favoring children high in passive responses.

APPENDIX C

ADDITIONAL NOTES ON SAMPLE AND METHODS

1969-70 Analysis

Analyses done on the 1969-70 PV data do not include all children and all sites. Individual children are excluded when the data recorded for them appears to be incomplete or invalid. In order to be included in the analyses, a child has to meet all of the following criteria:

1. The ages given for his pre- and post-test are comparable. This means that the chronological age listed for the spring test exceeds that given for the fall test by eight to ten months. Valid ages are crucial for the Binet, since the chronological age is used in computing IQ.
2. English is his first language.
3. His age at pre-test is 47 months or more.
4. His test scores are judged valid by the presiding tester.
5. Information on his age, sex, and ethnicity are complete.
6. His ethnicity is given as Black or white.

Three sites--Fort Walton Beach, Tuskegee, and Oraibi--are excluded from all 1969-70 IQ analyses. Fort Walton Beach and Tuskegee are eliminated because the Planned Variations study suspects the validity of 1969-70 IQ data from these sites. Oraibi is excluded from 1969-70 and 1970-71 analyses because all children in the site are American Indian; since no other site includes Indian children the sample can not be considered comparable.

1970-71 Analysis

Criteria for including individual children in the 1970-71 analysis are similar to those used for the 1969-70 data, with these additions:

a. Children are excluded if their age at pretest is less than 46 months.

b. Children are excluded if information on their prior preschool experience is missing.

c. Children are excluded from the smaller Stanford-Binet sample if either their PSI scores or their Stanford-Binet scores are missing or invalid.

The only site excluded from the 1970-71 analysis is Oraibi.

APPENDIX D

NOTES ON THE ANALYSIS OF THE 1970-71 PV DATA

This appendix describes the analyses of covariance and regression analyses performed on the 1970-71 sample.

Analyses of Covariance

I. Interactions relating to model

A. PSI sample - ANCOVA

Dependent variable: PSI post-test

Sample: all children in the eight models with valid scores on PSI pre- and post-test. N = 883.

Design: The preliminary analysis of PSI took in interactions of all categorical variables included in the 1969-70 analysis of post-test scores. The design was prior preschooling by sex by SES category by ethnicity by model, with covariates for age and pretest score. Because this analysis showed all interactions involving sex and SES category to be insignificant, the final model was prior preschooling by ethnicity by model, with covariates for SES, sex, age and pre-score.

B. IQ sample - ANCOVA

Dependent variables: PSI post-test, IQ post-test

Sample: all children in the eight models with valid scores on PSI and IQ pre- and post-tests. N = 305.

Design: The design of the analyses done on the Stanford-Binet

sample was model by "competence" by "passivity"[1] by sex by SES category by ethnicity;[2] covariates were age, IQ pre-score, and PSI pre-score.

II. General Interaction Study

Dependent variables: PSI and IQ post-tests; PSI and IQ gains

Sample: all children with valid scores on IQ and PSI pre- and post-test in Planned Variation or comparison classrooms. N = 607.

Design: ethnicity by preschool experience by SES category by "passivity" by "competence"; covariates included age, age as a dummy variable ("old" and "young" divided at the median age), PSI pre-score, IQ pre-score, and two-way interactions of these variables. Separate regression lines were allowed for where it was deemed necessary. The dependent variables were IQ and PSI post-test, and IQ and PSI gains (covariates on gains analyses were age and age as a dummy variable).

Regression Analyses

Unless otherwise noted, all regressions have been done stepwise with main effects forced in and interactions allowed to enter one by one to explain the maximum additional variance. Results given in the

---

[1]For explanation of these two Hertzig-Birch categories, see Part I, Section VII.

[2]Because only three of the eight models had more than eight children with prior preschool experience and valid pre- and post-test scores on both the PSI and the Stanford-Binet, interactions of preschooling with model could not be included in IQ analyses. Prior preschooling was eliminated as a covariate because for this sample it proved insignificant.

text are for the step on which the standard deviation of the residuals is minimum.

I. PSI regression 1[1]

Sample: all children with valid pre- and post-PSI scores in the eight models examined. N = 883.

a. Purpose: preliminary investigation of interactions of SES components with other variables.

Independent variables:[2] models, family size, mother's education, income, sex, ethnicity, preschooling, age, PSI pre-score, model by family size, model by mother's education, model by sex, model by race, model by preschooling.

b. Purpose: preliminary investigation of interactions of age and PSI pre-score with other variables.

Independent variables: models, family size, mother's education, income, sex, race, preschool experience, age, PSI pre-score; age by model, PSI pre-score by model, age by

---

[1]In regression analysis of this sort it is only possible to enter main effects for seven of the eight models. Under the assumption that model by variable interactions for the omitted model would be difficult to interpret, these interactions were also omitted. Thus, in order to find the equation which best described observed interactions, PSI model regressions were run twice, omitting the Gordon model on one run and the EDC model on another. The results were naturally very similar, except where an interaction with one of these two models was significant. Results are given for the run which include all significant interactions, but when the result for a model which is included on both runs is significant on one run and not on the other, this fact is noted in the text.

[2]Dummy variables used for models, sex, ethnicity, and prior preschooling.

preschool by model; all first-order interactions involving age and race.

c. Purpose: final equation describing interactions with model of significant SES variables, age and PSI pre-score.

Independent variables: models, mother's education, income, sex, ethnicity, preschooling, age, PSI pre-score; model by mother's education, model by income, model by sex, model by race, model by preschooling, model by age, model by PSI pre-score; model by age by preschooling; all first-order interactions involving race, preschooling, or age.

II. PSI regression 2

Sample: all children in the eight models with valid pre- and post-IQ and PSI scores. N = 305.

a. Purpose: to investigate model interactions with initial IQ and "competence".

Independent variables: model, mother's education, income, "competence", ethnicity, age, PSI pre-score, IQ pre-score, model by mother's education, model by income, model by competence, model by ethnicity, model by age, model by initial IQ; ethnicity by mother's education, income, "competence", age, and initial IQ; age by mother's education, income, competence and initial IQ.

III. PSI regression 3--grouped models

Sample: all children with valid pre- and post-PSI scores in Bushell, Engelmann-Becker, EDC, Bank Street, and Far West

programs. N = 422.

Purpose: investigation of interaction of all child variables with models grouped according to directiveness.

Independent variables: group (Engelmann-Becker and Bushell coded "more directive"; EDC, Bank Street, and Far West coded "less directive"), family size, mother's education, income, sex, ethnicity, age, preschooling, PSI pre-score; all first-order interactions involving group, ethnicity, preschooling, or age; all second-order interactions involving group and ethnicity, preschooling or age.

IV. PSI regression 4--grouped models

Sample: all children with valid pre- and post-PSI and IQ scores in Bushell, Engelmann-Becker, EDC, Bank Street, and Far West programs. N = 183.

a. Purpose: investigation of interaction of initial IQ, competence and prior preschooling with models grouped according to directiveness. N = 183.

Independent variables: group, family size, mother's education, income, "competence", "passivity", sex, ethnicity, preschooling, age, IQ and PSI pre-score; model group by family size, mother's education, income, "competence", "passivity", sex, ethnicity, preschooling, age, IQ pre-score; age by family size, mother's education, income, "competence", preschooling, sex, and PSI pre-score, ethnicity by family size, mother's education, income, "competence", preschooling, sex, PSI pre-score; preschooling

by family size, mother's education, income, competence, preschooling, sex, PSI pre-score; second-order interactions involving group and ethnicity, preschooling, or age; group by sex by IQ pre-score.

b. Purpose: investigation of interaction of PSI pre-score with groups.

Independent variables: as in PSI regression 4a, with group by PSI pre-score substituted for group by IQ pre-score interactions.

V. PSI regression 5

Sample: all children with valid pre- and post-PSI tests who have no prior preschool experience. N = 723.

Purpose: testing hypotheses relating to age by model interaction.

Independent variables: model, family size, mother's education, income, sex, ethnicity, age, PSI pre-score, model by family size, model by mother's education, model by income, model by sex, model by ethnicity, model by age, model by PSI pre-score, and all first-order interactions involving ethnicity or age.

VI. IQ regression 1

Sample: all children in the eight models with valid pre- and post-PSI and IQ scores. N = 305.

a. Purpose: investigation of main effects and model interactions with SES variables, age, and initial IQ.

Independent variables: model, family size, mother's education, income, "competence", "passivity", ethnicity, preschool experience, age, PSI pre-score, IQ pre-score; model by family size, model by income, model by mother's education, model by "competence", model by "passives", model by ethnicity, model by preschooling, model by age, model by initial IQ; ethnicity by family size, mother's education, income, "competence", "passives", preschool experience; age by family size, mother's education, income, preschooling, and initial IQ.

b. Purpose: further investigation of model interactions with initial IQ.

Independent variables: same as above.

Method: main effects and interactions of model with initial IQ forced in as far as possible. Other variables permitted to enter stepwise.

c. Purpose: to test hypotheses relating to interactions of SES components with model.

Independent variables: models, IQ and PSI pre-score, family size, mother's education, income, sex, ethnicity, preschooling, age, "competence", "passivity", ethnicity by age, ethnicity by mother's education, Tucson model by "passivity".[1]

Interaction of income, mother's education and family size with these models: Far West, Bank Street, Engelmann-Becker,

---

[1]These interactions are forced in because they enter significantly in other IQ analyses.

Bushell and Weikart.

Method: main effects, first three interactions forced in; interactions with SES components allowed to enter stepwise.

VII. IQ regression 2--grouped models

Sample: all children with valid pre- and post-IQ and PSI scores in Bushell, Engelmann-Becker, EDC, Bank Street and Far West programs. N = 183.

a. Purpose: investigation of interaction of all child variables with models grouped according to directiveness.

Independent variables: model-group, family size, mother's education, income, "competence", "passivity", sex, ethnicity, preschooling, age, PSI pre-score; group by family size, mother's education, income, "competence", "passivity", sex, ethnicity, preschooling, age, IQ pre-score; age by family size, mother's education, income, "competence", preschooling, sex, and PSI pre-score; ethnicity by family size, mother's education, income, "competence", preschooling, sex, PSI pre-score; preschooling by family size, mother's education, income, "competence", sex and PSI pre-score; second-order interactions involving group and ethnicity, preschooling, or age; group by sex by IQ pre-score.

b. Purpose: investigation of interaction of PSI pre-score with groups.

Independent variables: as in IQ regression 2a, with group by PSI pre-score substituted for group by IQ pre-score interactions.

VIII. IQ regression 3

Sample: all children in the eight models with valid pre- and post-IQ and PSI scores and no prior preschool experience. N = 239.

Purpose: testing hypotheses relating to age by model interactions.

a. Method: main effects and first three interactions force in. Interactions of age with Bushell and Tucson models forced in. Interactions of age with other models allowed to enter stepwise.

Independent variables: models, IQ and PSI pre-scores, family size, mother's education, income, "competence", "passivity", sex, ethnicity, age, interactions of age and mother's education with ethnicity, Tucson model by passivity,[1] model by age.

b. Method: main effects forced in; interactions permitted to enter stepwise.

Independent variables: model, age, ethnicity, sex of head of household, mother's education, income, family size, IQ pre-score; interaction of age with mother's education, sex of head of household, income, family size, sex, ethnicity, and model.

---

[1]These interactions are consistently significant on IQ analyses and are entered in order to improve the model.

APPENDIX E

CORRELATIONS AMONG INDEPENDENT AND DEPENDENT VARIABLES[1]
PLANNED VARIATIONS 1970-71

	age	sex	prior pre-schooling	ethnicity	mother's ed	income	family size	PSI pre-score	PSI post-score	IQ pre-score	IQ post-score	"competence"	"passivity"
age													
sex[2]	-.016												
prior preschooling[3]	-.233	.041											
ethnicity[4]	.120	-.045	-.052										
mother's ed.	-.197	-.069	-.053	.069									
income	-.128	-.083	-.028	.010	.158								
family size	-.069	-.027	-.064	.090	.201	-.051							
PSI pre-score	.444	.061	-.064	.265	.131	.120	.003						
PSI post-score	.420	.084	-.142	.283	.128	.099	.012	.773					
IQ pre-score	-.145	.075	-.205	-.186	.292	.174	-.166	.541	.526				
IQ post-score	-.315	.032	-.017	-.167	.254	.164	-.024	.299		.557			
"competence"[5]	.045	.054	.028	-.160	-.051	.076	.009	-.042	.001	-.076	.033		
"passivity"[5]	-.094	.025	-.049	.160	-.017	-.054	.135	-.104		-.152	-.073	.023	

[1] N = 883, except as noted below.

[2] Dummy variable: girls coded +1; boys coded -1.

[3] Dummy variable: children without prior preschooling coded +1; children with prior preschooling coded -1.

[4] Dummy variable: Blacks coded -1; whites coded +1.

[5] Dummy variables: children with one or more Hertzig-Birch responses in this category coded +1; children making no such responses coded -1.

APPENDIX E

SAMPLE SIZE FOR CORRELATIONS[1]

	age	sex	prior pre-school	ethnicity	mother's ed.	income	family size	PSI pre-score	PSI post-score	IQ pre-score	IQ post-score	"competence"	"passivity"
age	--												
sex		--											
prior pre-schooling			--										
ethnicity				--									
mother's ed.	764	764	764	764	--								
income	793	793	793	793	689	--							
family size	864	864	864	864	757	779	--						
PSI pre-score					764	793	864	--					
PSI post-score					764	793	864		--				
IQ pre-score	305	305	305	305	267	243	298			--			
IQ post-score	305	305	305	305	267	243	298			305	--		
"competence"	305	305	305	305	267	243	298			305	305	--	
"passivity"	305	305	305	305	267	243	298			305	305	305	--

[1] N = 883 where not otherwise noted.

BIBLIOGRAPHY

Anderson, G. L. Quantitative Thinking as Developed Under Connectionist and Field Theories of Learning. In E. J. Swenson et al., (Eds.), Learning Theory in School Situations. Minneapolis: University of Minnesota Press, 1949.

Anderson, G. L. A Comparison of the Outcomes of Instruction Under Two Theories of Learning. Unpublished doctoral dissertation, University of Minnesota, 1941.

Anderson, R. H., & Shane, H. G. As the Twig Is Bent. New York: Houghton Mifflin Co., 1971.

Bar-Yam, M. The Interaction of Instructional Strategies with Students' Characteristics. Harvard University, Center for Research and Development on Educational Differences, 1969.

Beller, E. K. The Evaluation of Effects of Early Educational Intervention on Intellectual and Social Development of Lower-Class Disadvantaged Children. In E. Grothberg (Ed.), Critical Issues in Research Related to Disadvantaged Children. Princeton: Educational Testing Service, 1969.

Bissell, J. S. The Cognitive Effects of Pre-school Programs for Disadvantaged Children. Unpublished doctoral dissertation, Harvard University, 1970.

Calvin, A. D., Hoffman, F. K., & Harden, E. L. The Effect of Intelligence and Social Atmosphere on Group Problem Solving Behavior. Journal of Social Psychology, 1957, 45, 61-74.

Crandall & Rabson. Children's Repetition Choices in an Intellectual Achievement Situation Following Success and Failure. Journal of Genetic Psychology, 97, 161-68.

Cronbach, L. J., & Snow, R. Individual Differences in Learning Ability as a Function of Instructional Variables. ERIC, March 1969.

Datta, L. A Report on Evaluation Studies of Project Head Start. Washington, D. C.: U. S. Department of Health, Education and Welfare, Project Head Start, 1969.

DiLorenzo, L., et al. Prekindergarten Programs for Educationally Disadvantaged Children. New York State Education Department, 1969.

Erickson, E. L., McMillan, J., Bennell, J., Hoffman, L., & Callahan, O.D. Experiments in Head Start and Early Education: The Effects of Teacher Attitude and Curriculum Structure on Preschool Disadvantaged Children. Washington, D. C.: Office of Economic Opportunity, Project Head Start, 1969.

Evans, J. W. Evaluating Social Action Programs. Washington, D. C.: Office of Economic Opportunity, 1969.

Gray, S., & Klaus, R. The Early Training Project. In Monographs of the Society for Research in Child Development, 33, 1968.

Head Start Planned Variation Study. Washington, D. C.: U. S. Department of Health, Education and Welfare, 1971.

Hertzig, M. E., Birch, G., Momas, A. & Mendez, O. A. Class and Ethnic Differences in the Responsiveness of Preschool Children to Cognitive Demands. Monographs of the Society for Research in Child Development, 33(1), 1968.

Hodges, W., Spicker, H., & McCandless, B. The Development and Evaluation of a Diagnostically Based Curriculum for Preschool Psycho-Socially Deprived Children. Washington, D. C.: U. S. Department of Health, Education and Welfare, 1964.

Jencks, C., Smith, M., Acland, H., Bane, M. J., Cohen, D., Gintis, H., Heyns, B., Michelson, S. Inequality. New York & London: Basic Books, 1972.

Kagan, J., & Moss, H. A. Birth to Maturity, A Study in Psychological Development. New York & London: John Wiley & Sons., Inc., 1962.

Karnes, M. B., et al. Research and Development Program on Preschool Disadvantaged Children, Vol. I. Washington, D. C.: U. S. Department of Health, Education and Welfare, 1969.

Kraft, I., Fuschillo, J., & Herzog, E. Prelude to School: An Evaluation of an Inner-city Program. Washington, D. C.: U. S. Department of Health, Education and Welfare, Children's Bureau, 1968.

Labor, W. The Logic of Non-Standard English. In Georgetown Monograph Series on Language and Linguistics, Monograph No. 22, 1969.

Lesser, G. Matching Instruction to Student Characteristics. In G. Lesser (Ed.), Psychology and Educational Practice. Glenview, Illinois: Scott Foresman & Co., 1971

Lukas, C. V. D., & Wohlleb, C. Implementation of Head Start Planned Variation, 1970-71. Cambridge, Mass.: Huron Institute, December 1972.

Maccoby, E. The Development of Sex Differences. Stanford, Calif.: Stanford University Press, 1966.

Maccoby, E., & Zellner, M. Experiments in Primary Education: Aspects of Project Follow-Through. New York: Harcourt Brace Jovanovich, Inc., 1970.

Miller, J. O. Review of Selected International Research with Young Children, ERK Clearinghouse on Early Childhood Education. Urbana, Illinois: University of Illinois, 1969.

Monaghan, A. Teachers in the Central School. Harvard Graduate School of Education, Unpublished paper, 1971.

Moriarty, A. E. Coping Patterns in Preschool Children in Response to Intellectual Test Demands. Genetic Psychology Monographs, 64, 3-167.

Nemenyi, P. Variances: An Elementary Proof and a Nearly Distribution-Free Test. The American Statistician, December 1969.

Palmer, Francis H. Minimal Intervention at Age Two and Three and Subsequent Intellective Changes, paper presented at Conceptualizations in Preschool Curricula Conference, May 1970.

Reissman, F. The Culturally Deprived Child. New York: Harper & Row, 1962.

Sarasen, S. B., Dandson, K., Wighthall, F. & Waite, R. Rorschach Behavior and Performance of High and Low Anxious Children. Child Development, 29(2), June 8, 1958.

Smith, M. P. Intellectual Differences in 5-year old Underprivileged Girls and Boys With and Without Pre-kindergarten School Experiences. Journal of Educational Research, 61, 1968, 348-350.

Smith, M. S. Some Short-term Effects of Project Head Start: A Preliminary Report on the Second Year of Planned Variation, 1970-71. Cambridge, Mass.: Huron Institute, January 1973.

Smith, M. S., & Bissell, J. S. Report Analysis: The Impact of Head Start. Harvard Education Review, 40, 51-104.

Soar, R. S. Follow-Through Classroom Process Measurement and Pupil Growth (1970-71). Institute for Development of Human Resources, University of Florida, Gainesville, Florida, June, 1972.

Soar, R. S. A Measurement of Teacher Classroom Management, Paper presented at a Symposium at the APA Meeting in Washington, D. C., September 1971.

Stanford Research Institute. Implementation of Planned Variation in Head Start: Preliminary Evaluations of Planned Variation in Head Start According to Follow-Through Approaches (1969-70). Interior Report on Contract HEW-OS-70-134. Menlo Park, California, March 1971.

Stearns, M. S. Report on Preschool Programs: The Effects of Preschool Programs on Disadvantaged Children and Their Families, Office of Child Development Report, draft, 1971.

Stodolsky, S. S. How Children Find Something to do in Preschools. Chicago: University of Chicago, 1972.

Van de Riet, V. A Sequential Approach to Early Childhood and Elementary Education, Phase III, Grant Report for Office of Child Development Grant No. OCD H8 222 B/H/O, University of Florida, Gainesville, Florida, 1972.

Walker, D. R., Bane, M. J., & Bryk, T. The Quality of the Head Start Planned Variation Data. Cambridge, Mass.: Huron Institute, 1973.

Ward, J. N. Group vs. Lecture-Demonstration Method in Physical Science Instruction. Journal or Experimental Education, 1956, 24, 197-210.

Weikart, D. A Comparative Study of Three Preschool Curricula. In J. Frost (Ed.), Disadvantaged Child (2nd ed.) New York: Houghton Mifflin, 1970.

Weikart, D. Relationship of Curriculum, Teaching, and Learning in Preschool Education. In J. C. Stanley (Ed.), Preschool Programs for the Disadvantaged: Five Experimental Approaches to Early Childhood Education. Baltimore and London: Johns Hopkins University Press, 1972.

Wispe, L. G. Evaluation Section Teaching Methods. Journal of Educational Research, 45(3), 1951.

Zigler, E., & Butterfield, E. Motivational Aspects of Changes in IQ Test Performance of Culturally-Deprived Nursery School Children, Child Development, 39(1), 1968.

Zimiles, H. Has Evaluation Failed Compensatory Education. In J. Hellmuth (Ed.), Disadvantaged Child, Vol. III, Brunner Mazel, 1970.